Indian Journal of Fundamental and Applied Life Sciences ISSN: 2231– 6345 (Online)
An Open Access, Online International Journal Available at www.cibtech.org/sp.ed/jls/2015/03/jls.htm
2015 Vol. 5 (S3), pp. 2291-2301/Rahmanpour et al.
Research Article

A SMOOTH CONSTRAINED PENALTY METHOD FOR OPTIMIZATION VIA RADIAL BASIS FUNCTIONS

[*]Fardin Rahmanpour[1], Mohammad Mehdi Hosseini[1, 2] and Farid Mohammad Maalek Ghaini[1]
[1]Department of Mathematics, Yazd University, Yazd, Iran
[2]Department of Applied Mathematics, Shahid Bahonar University, Kerman, Iran
**Author for Correspondence*

ABSTRACT
In this paper we have proposed a new approach for the penalization of the exact l_1 penalty function method. Then this method will be used for solving nonlinear constrained optimization problems with both equality and inequality constraints. However l_1 exact penalty functions are nonsmooth functions. In the proposed method a new penalty trust region based on Radial Basis Functions for smoothing the l_1 exact penalty function is introduced. Penalty function is replaced by a smooth surrogate model to analyze its global convergence. At each iteration, the trial step is determined such that either the value of the objective function or the measure of the constraint violation is sufficiently reduced. This method is particularly suitable for problems which contain a large number of constraints and decision variables. The numerical results are presented for some standard test problems.

Keywords: *Exact Penalty Function, Derivative-Free Method, Trust-Region Method, Non Smooth Optimization, Radial Basis Functions, Constrained Optimization, Penalized Optimization Problems*

INTRODUCTION
The class of nonlinear constrained optimization problems is an important class of problems with a broadrange in engineering, scientific, and operational applications. This paper introduces a new method which belongs to the class of trust-region methods via Radial Basis functions (RBFs) for solving the following constrained optimization problem:

$$
\begin{aligned}
&\min_x f(x)\\
&s.t.\\
&g_m(x) \leq 0 \qquad m \in I_1 = \{1,2,\dots,M\}\\
&h_l(x) = 0 \qquad l \in I_2 = \{1,2,\dots,L\}
\end{aligned}
\tag{1}
$$

where $x \in R^n$, $f: R^n \to R$ and $g_m (m \in I_1), h_l (l \in I_2): R^n \to R$ are not necessarily differentiable.
In general, one can not solve (1) directly or explicitly. Instead, an iterative method is used that solves a sequence of simpler, approximate subproblem to generate a approximate solution, $\{x_k\}$ of (1). The basic and classical constrained optimization methods include penalty function method and the Lagrangian method. The sequential quadratic programming method (SQP) is a well known and powerful constrained optimization method as well, which is a local search method able to find a local optimal solution.
We focus on using the traditional and effective quadratic penalty function framework as (Nocedal and Wright, 1999):

$$\min_x Q(x,\mu) = f(x) + \mu\{\textstyle\sum_{l=1}^{L} h_l^2(x) + \sum_{m=1}^{M} \max(g_m(x),0)\}, \tag{2}$$

For solving (1), where μ is the penalty parameter. Because (2) is not necessarily a smooth function, when constructing an algorithm to solve general problem (1) by (2), some difficulties may be faced.
Trust–region algorithms for the constrained optimization problems are a class of numerical algorithms for finding an approximate solution to problem (2). They are iterative methods that compute at every iteration k, a trial step by solving a trust–region subproblem. The step s_k is then tested using the merit function (2). The step s_k is accepted only if $x_{k+1} = x_k + s_k$ is a better approximation to the solution
The main difficulty associated with the penalty functions is the choice of the penalty parameter which can inordinate by increase. Another concern is the effect of the nondifferentiability of some popular exact penalty functions for which the penalty functions is allowed to increase for a limited number of iterations.

The basic and classical constrained optimization methods include penalty function method, the Lagrangian method (Luenberger and Ye, 2007) and SQP (Boggsal and Tolle, 1995). These methods have been developed under the assumption of uniqueness of the local optimal solution; that is, these are local search methods which can find a local optimal solution. Many constrained optimization methods that utilize meta–heuristics have been proposed in recent years (Coello, 2002; Lu and Chen, 2008**)**. Some are able to search for optimal solutions globally and have been used to successfully solve some constrained optimization problems (Liang and Suganthan, 2006; Takahama and Sakai, 2006) Generally, in meta–heuristics techniques, interaction among all search points is mainly used as the driving force. Therefore, these methods have the drawback that once all of the search points are attracted to one search point, diversity is lost and stagnation of the search occurs. The stagnation tends to occur in high dimensional optimization problems which have multi-peaked objective functions, and the meta-heuristics may not perform well in such problems. Zenios *et al.,* (1993) and Pinar and Zenios (1994) gave a smoooh exact penalty function for convex constrained optimization problems, which can be applied to obtain a good approximate optimal solution to (1). Nondifferentiable exact penalty functions were introduced for the first time by Eremin and Mazurov (1979) and Zangwill (1967). In almost all of the introduced penalized approaches the notion of convexity plays a dominant role.
Recently, derivative–free trust region algorithms have been used increasingly (Conn *et al.,* 1997; Kurokawa, 2009; Powell, 2002; Zhao *et al.,* 2006**)**. A common approach is to combine conventional algorithm such as genetic algorithm or pattern search with surrogate models to solve expensive problems. For instance, Booker *et al.,* (1998) and Jones *et al.,* (1998) 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 (1973) and reviewed by Powel (2002) and Conn *et al.,* (1997). Wild *et al.,* (2008**)** constructed a surrogate model based on RBFs.
In this study, we consider use of a trust–region method based on RBFs, which is able to obtain a global optimal solution without being trapped at global optimal solutions. The underlying idea of this method is that there are two goals, one is improving the feasibility and the other is reducing the value of objective function rapidly so that building a relation between reduction the value of objective function and improving the feasibility.
This study is aimed to minimize the penalty function (2) which is a non smooth function of several decision variables. Continuously differentiable functions with zero gradient is the standard condition for optimization techniques (Kolda *et al.,* 2003). But, here we can not use this theory exactly. Thus, derivative–free algorithms, which have a long history and are currently growing rapidly can be more effective because they extends only with value of function (Gray *et al.,* 2004; Regis and Shoemaker, 2007; Wild *et al.,* 2008; Zhao *et al.,* 2006).
In our approach, at each iteration, we construct the surrogate model by RBFs interpolation within a trust region instead of penalty function (2) to find the new trial minimum point. 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. This model is considered instead of the penalty function on a suitable trust region.
The main idea is to find interpolation points in the trust region for building a polynomial interpolation for objective function anyway that we can obtain strict local minimum point.
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 can be changed at each iteration. Under proper assumptions, the proposed method will guarantee global convergence. Also, models based on RBFs have been shown to be of interest for global optimization.
The 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, $f_k = f(x_k)$, $g_m = g(x_m)$ and $h_l = h(x_l)$.

Surrogate Model

We consider the following unconstrained optimization problem

$$\min_{x \in R^n} Q(x, \mu),$$

where $Q: R^n \to R$ is a merit function which is necessarily non smooth. In this paper, we propose the smooth surrogate model, which is minimized instead of $Q(x, \mu)$ more easily.

We remark that surrogate modeling 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.

Quadratic Surrogate Model

Powell (2002) and Conn *et al.*, (1996, 1997) proposed the surrogate model as follows:

$$Sm(x_k + s, \mu) = Q_k + G_k^T s + \frac{1}{2} s^T H_k s,$$

where $G_k = \nabla Q(x_k, \mu)$ and $H_k = \nabla^2 Q(x_k, \mu)$. When Q is twice differentiable and admits a Hessian matrix H which will always be positive definite.

The goal is to construct the surrogate model $Sm(x_k, \mu)$ for the objective function Q, 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 $B_k := \{x \in R^n : ||x - x_k|| \leq \Delta_k\}$. At each iteration of the surrogate method, the solution of optimization problem inside B_k (Nelder and Mead, 1965; Conn *et al.*, 2000), as

$$\min_s Sm(x_k, \mu) s.t. ||s|| \leq \Delta_k, \tag{3}$$

is needed, for some trust region with radius $\Delta_k \geq 0$. We compute $Q(x_k + s, \mu)$, and define:

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

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

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

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

and

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

The following assumptions were considered in this section:

1. $Q(x, \mu)$ is a two times differentiable function.
2. $\{x_k\}$ is a bounded sequence.

Suppose the these assumption holds. Let s_k be a solution of subproblem (3). The following lemma, which can be obtained from the well-known result is needed (Powell; 2002).

Lemma 2.1 *subproblem (3) has a sufficient decrease condition if,*

$$Sm_k(x_k, \mu) - Sm_k(x_k + s_k, \mu) \geq \frac{c}{2} ||G_k|| \min(\Delta_k, \frac{||G_k||}{||H_k||}),$$

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

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

Radial Basis Functions Interpolation

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

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

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

where $\varphi(||s - y^i||)$ is the RBFs centered at the point y^i. Note that we have $V(s) = \sum_{j=1}^{M} \gamma_j \nu_j(s)$, where $\nu = \{\nu_1(s), \dots, \nu_M(s)\}$ is an order basis for the linear space Π^n_{M-1}, the space of polynomials of total degree less than or equal to $M - 1$, with n variables and $\{\lambda_j\}_{j=1}^{N}$ are the unknown RBF coefficients. In the following conditions, the approximation $Sm(x, \mu)$,

$$Sm(x_i, \mu_i) = Q_i, \qquad i = 1, \dots, N, \tag{5}$$

$$\sum_{i=1}^{N} \lambda_i \nu_k(s_i) = 0, \qquad k = 1, \dots, M, \tag{6}$$

the conditions (5) and (6) can be written as matrix,

$$\begin{bmatrix} \Phi & V \\ V^T & \varnothing \end{bmatrix} \begin{bmatrix} \lambda \\ \nu \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}. \tag{7}$$

Which coefficients λ is the undetermined coefficient vector. The sake of clarity, the matrix Φ 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 (7) is well-posed if the coefficient matrix is non-singular (Buhmann, 2003). Micchell (1986) proved that the interpolation problem in equation (7) is solvable when the following two conditions are met:

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

Definition 3.1 *Let ν be a basis for Π^n_{M-1}, with the convention that $\nu = \emptyset$ if $M = 0$. A function ϕ is said to be conditionally positive definite (CPD) of order M if for all the distinct points $Y \subset R^n$ and all $\lambda \neq 0$, satisfying $\sum_{i=1}^{N} \lambda_i \nu(x_i) = 0$, the quadratic form $\sum_{i,j=1}^{N} \lambda_j \phi(||x_i - x_j||)\lambda_j$ is positive* (Buhmann, 2003; Wendland, 2005).

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

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

For fixed coefficients λ, these radial basis functions are all twice continuously differentiable. Therefore we have relatively simple analytic expressions for the gradient:

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

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

Derivative Optimization

In this section, we present an algorithm for reducing penalty function of several variables. Derivative-free optimization method has a long history and we refer the reader to (Conn *et al.*, 1997) for further references.

We suppose that $Q(x, \mu)$ is a function from R^n into R which is not necessarily smooth. The algorithm is based on approximating the penalty function by a positive definite quadratic model. The main idea is to use the available values of the penalty function and building a quadratic model by interpolating within a trust region.

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

$$Sm_k(x_k + s, \mu) = Q_k + G_k^T s + \frac{1}{2} s^T H_k s. \qquad (8)$$

Where the vector $G_k \in R^n$ and $H_k \in R^{n \times n}$ be a symmetric matrix. By imposing the interpolation condition in what follows:

$$Sm_k(x + y^i, \mu) = Q(x + y^i, \mu), \qquad i = 1, \dots, N. \qquad (9)$$

It is needed to evaluate $Sm_k(x + s)$ on $N = \frac{1}{2}(n + 1)(n + 2)$ points to find an approximating quadratic form, where n is number of variables (Andrew *et al.*, 2009; Bjorkman and Holmstrom, 2000; Conn *et al.*, 1997).

We consider $\{\varphi_i(.)\}_{i=1}^N$ as a basis for the linear space of n-dimensional quadratic function. The quadratic function (8) can be expressed as,

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

for some coefficients λ_i that could be determined from the interpolation condition (9),

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

λ_i are unique if the determinant of the below matrix is nonzero

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

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

Surrogate Methods Based on Radial Basis

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

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

Model is twice differentiable and is important for the convergence part of our method (Conn, Toint, 1996; Qeuvray, 2005). This study considers interpolation condition at the points of Y:

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

Indian Journal of Fundamental and Applied Life Sciences ISSN: 2231– 6345 (Online)
An Open Access, Online International Journal Available at www.cibtech.org/sp.ed/jls/2015/03/jls.htm
2015 Vol. 5 (S3), pp. 2291-2301/Rahmanpour et al.
Research Article

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

$$\begin{bmatrix} \Phi & V \\ V^T & \varnothing \end{bmatrix}\begin{bmatrix} \Lambda \\ \Gamma \end{bmatrix} = \begin{bmatrix} Q \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} \Phi & V \\ \varnothing & -V^T\Phi^{-1}V \end{bmatrix}\begin{bmatrix} \Lambda \\ \Gamma \end{bmatrix} = \begin{bmatrix} Q \\ -V^T\Phi^{-1}Q \end{bmatrix}, \tag{10}$$

with the solution $\Gamma = (V^T\Phi^{-1}V)^{-1}V^T\Phi^{-1}Q$, $\Lambda = \Phi^{-1}(Q - V\Gamma)$.
Sufficient condition for the solvability of system (10) is that the points in Y are distinct and yield a V^T of full column rank.
Suppose that $V^T = QR$ and hence $R \in R^{(n+1)\times(n+1)}$. If Z is an orthonormal basis for the null space of V (Benzi *et al.*, 2005), using the condition (6), it follows that $\Lambda \in \aleph(V)$. Therefore, $\Lambda = Zw$. According to (10), $\Phi\Lambda + V^T\Gamma = Q$. Multiplying this equation by Z^T from left gives, $Z^T\Phi\Lambda + Z^TV^T\Gamma = Z^TQ$. Keeping in mind that Z is an orthonormal basis for the null space V, we obtain $Z^TV^T\Gamma = 0$. Hence

$$Z^T\Phi Zw = Z^TQ. \tag{11}$$

Now, we can obtain w from (11) and thus we can compute the vector Λ By introducing the RBFs based on cubic spline (Andrew *et al.*, 2009; Buhmann, 2003) 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 (11), $LL^Tw = Z^TQ \Rightarrow w = (LL^T)^{-1}Z^TQ$, so that

$$||\Lambda|| = ||Zw|| = ||ZL^{T^{-1}}L^{-1}Z^TQ|| \le ||L^{-1}||^2|Q|,$$

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

$$R\Gamma = Q^T(Q - \Phi Zw). \tag{12}$$

In this section, we discuss a method of creating surrogate models. For this purpose Φ must be conditionally positive definite of order at least 2 (table1), and $V \in \Pi_2^n$ be linear. The RBFs interpolation is defined such that at all sample points the equation is established. Note that Φ must be conditionally positive definite of order at least 2 (table1) (Wild *et al.*, 2008).
The RBF coefficients λ_i and v_i must be bounded in magnitude. Define y^i to be ith point in Y, that is in the vicinity of the trust region. However for $n \ge 1$ condition (9) is not sufficient for the existence and uniqueness of the interpolant, 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 interpolant (Conn *et al.*, 2008).
The process can be summarized as follows: the study has chosen sample points in the vicinity of the trust region so that $n + 1$ offinely independent points and generate the other interpolation points.
The cubic spline $\varphi(r) = r^3$ in dimension n is unisolvent on points $Y = \{y^1, \dots, y^N\}$ if the matrix,

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

is invertible for any choice of N distinct points $y^1, \dots, y^N \in Y$.
Definition 5.1 *Y is unisolvent for Π_M^n if there exists a unique polynomial in π_M^n 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 (10) which is non-singular if the interpolation point set Y is unisolvent.
The collection of $n + 1$ distinct points will uniquely determine a polynomial of lowest possible degree in Π^n. In this section we describe an algorithm to find $n + 1$ interpolation points which are offinelly independent points. We denote $D := \{d_i \in \Delta_k | Q(x_k + d_i) is known\}$. Algorithm 5.1 shows that how to obtain $n + 1$ offinelly interpolation points.
Algorithm 5.1 *Finding n+1 offinely independent points:*
Step0: Input Y, constants $0 < \gamma_0 \le \gamma_1$, $\Delta_k \in [0, \Delta_{max}]$.
Step1: Define $D = \{d_1, d_2, \dots, d_{|D|}\} \in R^n$ such that $x_i = x_k + d_i$ are close to x_k.

Step2: We consider Z as I_n.
Iteration $k \geq 1$
Step3: Consider $d_i \in D$
If $||d_i|| \leq \gamma_1.\Delta_k$, define $u = \frac{d_i}{\gamma_0 \Delta_k}$,
If $||proj_Z^u|| \geq \gamma_0$, $\{d_i\} \in Y$,
Using the Gram-Schmidt, we obtain orthonormal basis for Y as $\bar{Z}$, update $Z = \bar{Z}$.
If $|Y| = n + 1$ stop.
Increment i by one, and go to Step 3.
Step4: If $|Y| < n + 1$, choose $d_i \in D$
If $||d_i|| \leq 2\Delta_{max}$, define $u = \frac{d_i}{\gamma_0 \Delta_k}$
If $||proj_Z^u|| \geq \gamma_0$, set $\{d_i\} \in Y$
Using the Gram-Schmidt process, we obtain an orthonormal basis for Y as $\bar{Z}$
Step5: Update $Z = \bar{Z}$ Increment i by one, and go to Step 4.

But when the number of points is $n + 1$, solution of the system (10) is just interpolation for linear functions and coefficient $\Lambda = \emptyset$. To build the surrogate model for nonlinear functions, we add some new points. Algorithm 5.2 shows how we can obtain "well independent" additional sample points in the trust region.

Algorithm 5.2 *Finding additional independent points:*
Step0: Input Y, $p_{max} = \frac{(n+1)(n+2)}{2}$, $D = \{d_1, d_2, \dots, d_{|D|}\}$, $\theta > 1$.
Iteration $k \geq 1$
Step1: Consider $d_i \in D$

$$\Pi^T = \begin{bmatrix} y^1 = 0 & y^2 & \dots & y^{|Y|} & d_i \\ 1 & 1 & \dots & 1 & 1 \end{bmatrix}, d_i \in D$$

Step2: Find the orthogonal basis Z for null space Π .
Step3: Build the interpolation matrix by using the cubic spline function at sample points Y,

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

Step4: Multiplying Φ_{new} by Z and Z^T, concludes,

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

Step5: P is positive definite for cubic spline function $\varphi(r) = r^3$, note that if P must be positive definite, the points Y must be distinct.
Step6: $P = LL^T$, if diagonal entries of L^T are positive, d_j add to sample points Y. If $|Y| = p_{max}$ Stop.
Increment i by one, and go to Step 1.
Step7: If there are no adaptable point in Y to independent, enlarged region as $\Delta_k = \theta.\Delta_k$, and go to Step 0.

Consider $Z^T \Phi_{new} Z = LL^T$, if all diagonal entries of L^T are positive, then d_j is added to the interpolation set Y. This procedure continues until $|Y| = \frac{(n+1)(n+2)}{2}$. Also the points $D = \{d_1, \dots, d_{|D|}\}$ are smartly chosen around the trial point x_k by a random process.

Optimization Surrogate on Radial Basis Functions (OSRB)

This section discusses the details of the OSRB algorithm (Abramson *et al.*, 2008**;** Forrester and Keane, 2009**)**. This paper is different from previous ones, by using a cubic spline function constructed via the surrogate model sake optimization.

Given N, a set of distinct interpolation points $Y = \{y^1 = 0, y^2, \ldots, y^N\}$, which $y^i \in R^n, i = 1 \ldots N$, we obtain the surrogate model for Q on Y. Algorithm 6.1 shows how surrogate model is constructed by RBFs.

Algorithm 6.1 *Iteration k of a derivative free surrogate model:*
Step0: Input $0 \le \eta_0 < \eta_1 \le 1$, $\gamma_0 < 1 < \gamma_1$, $0 < \Delta_1 \le \Delta_{max}$, $\varepsilon > 0$ and $\mu > 0$.
We assume that trial point x_k is given.
Iteration $k \ge 1$
Step1: From algorithm 5.1 and 5.2 find independent points that is denoted by Y.
Step2: Obtain surrogate model $Sm_k(x_k + s, \mu)$ by using the RBF's described in Section 5.
Step3: While $||\nabla Sm_k|| > \varepsilon$
3.1 Obtain a step s_k by solving: $\min\{Sm_k(x_k + s, \mu_k); x_k + s \in B(x_k, \Delta_k)\}$.
3.2 Evaluate $Q(x_k + s_k, \mu_k)$ and update the trial point according to the ratio ρ_k,

$$x_{k+1} = \begin{cases} x_k + s_k & \rho_k \ge \eta_1 \\ x_k + s_k & \rho_k > \eta_0 \\ x_k & o.w, \end{cases}$$

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

Step4: Choose new penalty parameter, $\mu_{k+1} > \mu_k$ and $x_{k+1} = x_k$;
In step 1, the interpolation point set $Y = \{y^1 = 0, \ldots, y^{|Y|}\}$ is determinate and is linearly independent. In step 2, we consider how to construct a model and to obtain parameters of RBFs model from (11), and (12). In step 3, the algorithm has criteria for model $Sm_k(x_k + s, \mu_k)$ and updates parameters trust region method. We finds the candidate step s_k by approximately solving the subproblem (3). In this paper, we solve subproblem (3) by using the Fmincon in Matlab.

Convergence Properties of OSRB Algorithm
In this section, we discuss the convergence properties of the algorithm. The trust region algorithm ensures the penalty function $Q(x, \mu)$ is approximated within a suitable neighborhood of x,
$L(x_0) := \{y \in R^n \,|\, ||x - y|| \le \Delta_{\max}, \forall x; Q(x, \mu) \le Q(x_0, \mu)\}$,
Assumption 6.1 *The penalty function $Q(x, \mu)$ is bounded below on $L(x_0)$ and $Sm(x, \mu)$ is twice continuously differentiable.*
Theorem 6.1 *Let $\{\Delta_k\}$ and $\{x_k\}$ be sequences generated by OSRB Algorithm. Then,* $\lim_{k \to \infty} \Delta_k = 0$.
Proof. After the last successful iteration, there is an infinite number of iterations that are either acceptable or unsuccessful, and in either case the trust region is reduced. If $x_{k+1} = x_k + s_k$ is obtained so that $Q(x_{k+1}, \mu) \le Q(x_k, \mu)$, then Δ_k is never increased for sufficiently large k, so Δ_k is decreased at least once every n iterations by a factor of $0 < \gamma < 1$, thus Δ_k convergence to zero.
The statement of theorem 6.1 gives a natural stopping criterion for OSRB algorithm. It results from the updating of the trust region at the k iteration. Surrogate model Sm_k is made so that,

$$Sm(s_k, \mu_k) - Sm(0, \mu_k) = G(0)^T s_k + \frac{1}{2} {s_k}^T H_k(0) s_k, \qquad \blacksquare$$

where $G_k = \nabla Q(x_k, \mu)$ and $H_k = \nabla^2 Q(x_k, \mu)$.
By Assumption 6.1 there exists $M > 0$ such that $||\nabla^2 Sm_k|| < M$, based on Assumption 6.1, we have the following lemma.
Lemma 6.1 *Suppose that assumption 6.1 holds. Then,*

$$Sm(0, \mu_k) - Sm(s_k, \mu_k) \ge \frac{1}{2} ||G_k|| \min\{\Delta_k, \frac{||G_k||}{H_k}\}.$$

Proof. If $|s_k| = ||-\frac{G_k}{H_k}|| \le \Delta_k$, then the quadratic subproblem (8) can be resolved,

$$Sm(s_k, \mu_k) = Sm(0, \mu_k) - \frac{G_k}{H_k} G_k + \frac{1}{2} (-\frac{G_k}{H_k})^T H_k (-\frac{G_k}{H_k}),$$

given the cubic spline is twice continuously differentiable, $G_k^T H_k G_k$ is positive definite. We know the model is convex along direction s_k. Next,

$$Sm(0,\mu_k) - Sm(s_k,\mu_k) = \frac{||G_k||^2}{H_k} - \frac{1}{2}\frac{||G_k||^2}{H_k} \geq \frac{1}{2}||G_k||\min\{\Delta_k, \frac{||G_k||}{||H||_k}\},$$

lemma 6.1 guarantees that the OSRB Algorithm will sufficient decrease at iteration k.

Numerical Results

Here, the proposed model is demonstrated for solving a numerical example. Consider the following optimization problem:

$$\min_x f(x) = -0.25x_1^2 + 1.2x_2^2 + 0.1x_1x_2 + 2x_1 - 2x_2;$$
$$s.t.$$
$$g_1(x) = x_1 - x_2 + 2 \geq 0$$
$$g_2(x) = -x_1 + 2x_2 + 2 \geq 0$$
$$g_3(x) = x_1x_2 \geq 0$$
$$h_1(x) = x_1 + 2x_2 - 3$$

As $f(x)$ and $g_3(x)$ are nonconvex. This problem has a optima point $x^* = (0,1.5)^T$. The best optimal solution using the proposed method is equal to $(2.1087399153e - 12, 1.499957965)$, which seems to be the same as the optimal value.

We present a set of nonlinear programming problems from (Hock and Schittkowski, 1981; Schittkowski, 1987) which have been solved by the OSRB algorithm proposed to accommodate practical experiment to show the success of proposed method. Notice that the interpolation points are choosed so that interpolation matrix (10) always 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 the trust region. We solve problem (1) with equality and inequality constraints to show the efficiency of the proposed method.

For the nonlinear programming problems we use the following OSRB parameters: $\Delta_1 = \max(1, ||x_0||)$, $\Delta_{\max} = 10^{-3}\Delta_1$, $\kappa_f = 10^{-10}$, $\kappa_d = 10^{-10}$, $\eta_0 = 0$, $\eta_1 = 10^{-3}$, $\gamma_0 = 0.1$, $\gamma_1 = 10$, and termination criterion $||\nabla Sm(x_k, \mu)|| < 1.e - 7$.

The numerical results for the test problems are listed in table 2. The header of the columns mean that: n and m are number of variables and constraints respectively, Δ_k is the radius of final trust region and f_{OSRB} is the final value of the objective function value at the final iteration.

Table 2: Numerical results

Problem	n	m	Δ_k	f_{OSRB}	**Problem**	n	m	Δ_k	f_{OSRB}
HS1	2	1	2.23E-9	1.42E-12	**HS30**	2	7	10	1
HS2	2	1	1.13E-19	5.04267	**HS32**	3	5	1E-7	1
HS3	2	1	1.4E-8	1.1E-14	**HS53**	5	9	.47E-8	4.0929
HS4	2	2	8.82E-7	2.66666714	**HS60**	3	4	3.2E-6	0.32567
HS6	2	1	1.83E-7	7.78E-17	**S217**	2	3	1E-9	-0.8
HS10	2	1	1E-8	-1	**S226**	2	4	1E-7	-0.49998
HS11	2	1	8.6E-8	-8.49851	**S227**	2	2	1.33E-08	1
HS12	2	1	1E-8	-30	**S228**	2	2	2.99E-4	-3
HS14	2	2	2.82E-7	1.3934650	**S230**	2	2	1.91E-09	0.375
HS16	2	5	2.2E-9	0.25	**S234**	2	5	1E-4	-8
HS17	2	5	2.23E-9	1	**S248**	3	2	1E-8	-0.79998
HS20	2	5	2.1E-7	38.1987	**S249**	3	1	1.73E-5	1
HS21	2	5	1.41E-8	-99.959	**S262**	4	8	2E-3	-9.9998
HS22	2	2	2.82E-7	1	**S263**	4	4	1.73E-5	-1
HS23	2	7	6.3E-9	2	**S264**	4	3	3.5E-7	-44

In Table 2, the problems are numbered in the same way as in Hock and Schittkowski **(**Hock and Schittkowski, 1981) and Schittkowski (Schittkowski, 1987). For example, HS6 means problem 6 in Hock and Schittkowski collection **(**Hock and Schittkowski, 1981) and S248 means problem 248 in Schittktwski collection (Schittkowski, 1987). We present some classical numerical examples which are solved by using the proposed algorithm in table 2.

CONCLUSION

For the l_1 exact penalty function with a large number of equality and inequality nonlinear constrained optimization, we propose a new smoothing algorithm based on the trust-region method by RBFs, and discuss global convergence. At each iteration, a surrogate model is constructed instead of the penalty function by RBFs. The most advantage of the proposed algorithm is that the management of the interpolation points is easier, so that the system (10) always has a unique solution. It focuses on improving feasibility or reducing of the objective function, while the value of the penalty parameter is not overally increased. We have tested a set of problems from (Hock and Schittkowski, 1981; Schittkowski, 1987. The preliminary numerical results show that the new method is effective. Future work perspective will be concerned with the algorithm performance. More numerical experiments, especially for large scale problems, should be done. The idea of the new algorithm is worthy to use free-penalty method of constrained optimization problems by this method in the future.

ACKNOWLEDGEMENT

This research was financially supported by the Center of Excellence for Robust and Intelligent Systems of Yazd University. We are grateful to Yazd University, for their useful collaboration.

REFERENCES

Abramson MA, Asaki TJ, Dennis Jr JE, OReilly KR and Pingel RL (2008). Quantitative object reconstruction via Abel-based X-ray tomography and mixed variable optimization. *SIAM Journal on Imaging Sciences* **1**(3) 322–342.

Andrew R, Conn AR, Scheinberg K and Vicente LN (2009). Introduction to Derivative-Free Optimization. *SIAM* **8**.

Benzi M, Golub GH and Liesen J (2005). Numerical solution of saddle point problems. *Acta Numerica* **14**(1) 1–137.

Bjorkman M and Holmstrom K (2000). Global Optimization of Costly Nonconvex Functions Using Radial Basis Functions. *Optimization and Engineering* **1**(4) 373–397.

Boggsal Paul T and Tolle Jon W (1995). Sequential quadratic programming. *Acta Numerica* **4** 1-51.

Booker AJ, Frank PD, Dennis Jr, Moore DW and Serani DB (1998). Managing surrogate objectives to optimize a helicopter rotor design. *Further Experiments AIAA MDO* 98-4717.

Buhmann MD (2003). *Radial Basis Functions: Theory and Implementations* (Cambridge University Press) Cambridge, England **12**.

Coello CA (2002). Theoretical and numerical constraint–handling techniques used with evolutionary algorithms: A survey of the state of the art. *Computer Methods in Applied Mechanics and Engineering* **191**(11) 1245-1287.

Conn AR and Toint PHL (1996). An algorithm using quadratic interpolation for unconstrained derivative free optimization. Nonlinear Optimization and Applications, Springer US 27-47.

Conn AR, Guold M and TOINT PL (2000). *Trust-Region Method* (MPS-SIAM Series on Optimization SIAM, Philadelfia) **1**.

Conn AR, Scheinberg K and Toint PHL (1996). On the convergence of derivative-free methods for unconstrained optimization. In: *Approximation Theory and Optimization: Tributes to MJD Powell*, edited by Buhmann MD and Iserles A (Cambridge University Press) 83-108.

Conn AR, Scheinberg K and Toint PHL (1997). Recent progress in unconstrained nonlinear optimization without derivatives, *Mathematical Programming* **79**(1-3) 397–414.

Conn AR, Scheinberg K and Vicente LN (2008). Geometry of interpolation sets in derivative free optimization. *Mathematical Programming* **111**(1-2) 141–172.
Eremin II and Mazurov VD (1979). Nonstationary Processes of Mathematical Programming (Nestatsionarnye protsessy matematicheskogo programmirovaniya).
Forrester AIJ and Keane AJ (2009). Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences* **45**(1) 50-79.
Gray G, Kolda T, Sale K and Young M (2004). Optimizing an empirical scoring function for transmembrane protein structure determination. *INFORMS Journal on Computing* **16**(4) 406–418.
Hock W and Schittkowski K (1981). Test examples for nonlinear programming codes, Lecture Notes in Economics and Mathematics Systems, Springer-Verlag, Berlin, **187**.
Jones DR, Schonlau M and Welch WJ (1998). Efficient global optimization of expensive black–box functions. *Journal of Global Optimization* **13**(4) 455–492.
Kolda TG, Lewis RM and Torczon VJ (2003). Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review* **45**(3) 385–482.
Kurokawa N (2009). Global convergence of the derivative free trust region Algorithm using inexact information a function values. PhD thesis, Graduate School Of Informatics Kyoto University.
Liang JJ and Suganthan PN (2006). Dynamic multi-swarm particle swarm optimizer with a novel constraint–handling mechanism. In *Proceeings of IEEE Congress on Evolutionary Computation* 9-16.
Lu H and Chen W (2008). Self–adaptive velocity particle swarm optimization for solving constrained optimization problems. *Journal of Global Optimization* **41**(3) 427–445.
Luenberger DG and Ye Y (2007). *Linear and Nonlinear Programming*, 3rd edition (Springer Science & Business Media) **116**.
Nelder JA and Mead R (1965). A simplex method for function minimization. *The Computer Journal* **7**(4) 308–313.
Nocedal JA and Wright SJ (1999). *Numerical Optimization* (Springer Science & Business Media, New York).
Pinar MC and Zenios SA (1994). On Smoothing Exact Penalty Functions for Convex Constrained Optimization. *SIAM Journal on Optimization* **4**(3) 486-511.
Powell MJD (2002). UOBYQA: unconstrained optimization by quadratic approximation. *Mathematical Programming* **92**(3) 555–582.
Qeuvray R (2005). Trust-Region Methods Based on Radial Basis Functions with Application to Biomedical Imaging. PhD thesis, Ecole Polytechnique Fdrale de Lausanne.
Regis RG and Shoemaker CA (2007). A stochastic radial basis function method for the global optimization of expensive functions. *INFORMS Journal on Computing* **19**(4) 497-509.
Schittkowski K (1987). *More Test Examples for Nonlinear Programming Codes* Lecture Notes in Economics and Mathematical Systems (Springer-Verlag, Berlin, Heidelberg) **282**.
Takahama T and Sakai S (2006). Constrained optimization by the Constrained Optimization by the Constrained Differential Evolution with Gradient–Based Mutation and Feasible Elites. In *Proeedings of IEEE Congress on Evolutionary Computation* 1–8.
Wendland H (2005). *Scattered Data Approximation* (Cambridge University Press) Cambridge, England.
Wild SM, Regis RG and Shoemaker CA (2008). ORBIT: Optimization by radial basis function interpolation in trust-regions. *SIAM Journal on Scientific Computing* **30**(6) 3197-3219.
Winfield D (1973). Function minimization by interpolation in a data table. *SIMA Journal of Applied Mathematics* **12**(3) 339–347.
Zangwill WI (1967). Nonlinear Programming via Penalty Function. *Manangement Science* **13**(5) 334 358.
Zenios SA, Pinar MC and Dembo RS (1995). A Smooth Exact Penalty Function Algorithm for Network–structured Problems. *European Journal of Operations Research* **64**(1) 258-277.
Zhao Z, Meza JC and Van Hove M (2006). Using pattern search methods for surface structure determination of nanomaterials. *Journal of Physics: Condensed Matter* **18**(39) 8693–8706.