RF cavity design exploiting a new derivative-free trust region optimization approach

Abdel-Karim S.O. Hassan a, Hany L. Abdel-Malek a, Ahmed S.A. Mohamed a,*, Tamer M. Abulfadel b, Ahmed E. Elqenawy a

a Engineering Mathematics and Physics Department, Faculty of Engineering, Cairo University, Giza 12613, Egypt
b Electronics and Electrical Communications Department, Faculty of Engineering, Cairo University, Giza 12613, Egypt

ARTICLE INFO

Article history:
Received 3 April 2014
Received in revised form 18 August 2014
Accepted 20 August 2014
Available online 30 August 2014

Keywords:
Optimal design
Derivative-free optimization
Trust region
Quadratic surrogate model
Linear accelerator

ABSTRACT

In this article, a novel derivative-free (DF) surrogate-based trust region optimization approach is proposed. In the proposed approach, quadratic surrogate models are constructed and successively updated. The generated surrogate model is then optimized instead of the underlined objective function over trust regions. Truncated conjugate gradients are employed to find the optimal point within each trust region. The approach constructs the initial quadratic surrogate model using few data points of order \( O(n) \), where \( n \) is the number of design variables. The proposed approach adopts weighted least squares fitting for updating the surrogate model instead of interpolation which is commonly used in DF optimization. This makes the approach more suitable for stochastic optimization and for functions subject to numerical error. The weights are assigned to give more emphasis to points close to the current center point. The accuracy and efficiency of the proposed approach are demonstrated by applying it to a set of classical bench-mark test problems. It is also employed to find the optimal design of RF cavity linear accelerator with a comparison analysis with a recent optimization technique.

© 2014 Production and hosting by Elsevier B.V. on behalf of Cairo University.

Introduction

In general, engineering systems are characterized by some designable parameters and some performance measures. The desired performance of a system (design specifications) is described by specifying bounds on the performance measures of the system which is set by the designer. The conventional system design aims at finding values of the system designable parameters that merely satisfy the design specifications. In general, there will be a multitude of acceptable designs. However, for contemporary engineering design, other criterion (objective function) can be chosen for comparing the different alternative acceptable designs (optimization problem) and for selecting the best one (optimal system design). Naturally, system performance measures and the objective functions are functions of system parameters values and evaluated through system simulations. For intensive CPU engineering systems, the high expense of the required system simulations may obstruct the optimization process.
In practice, robust optimization methods that utilize the fewest possible number of function evaluations are greatly required [1,2]. Another difficulty is the absence of any gradient information as the required simulations cost in evaluating the gradient information is prohibitive in practice [3]. Attempting to approximate the function gradients using the finite difference approach requires much more function evaluations, which highly increase the computational cost. Another objection in estimating the gradients by finite differencing is that the estimated function values are usually contaminated by some numerical noise due to estimation uncertainty. Hence, gradient-based optimization methods cannot be applied here.

For such optimization problems, only derivative free optimization (DFO) methods can be applicable [1,4–7]. Further, the derivative free trust region methods usually handle such problems more efficiently as the trust region framework constitutes one of the most important globally convergent optimization methods, which has the ability to converge to a solution starting from any arbitrary initial point [8]. In addition, these methods use computationally cheap surrogate-based models that can be constructed by using function evaluations at some selected points. These surrogate models may be response surfaces, radial basis method, neural networks, kriging, etc. The majority of the existing derivative-free trust region techniques have the following features:

- They require a relatively large number of function evaluations, $O(n^2)$ (where $n$ is the number of system design variables) to construct the initial quadratic model.
- The quadratic surrogate models are constructed via interpolating the objective function at a constant number of points; when a point is obtained a previous point is dropped. In addition, these algorithms usually ignore the valuable information contained in all previously evaluated expensive function values.

The work presented in this article introduces a new derivative free trust region approach that neither require nor approximate the gradients of the objective function. It implements a non-derivative optimization method that combine a trust region framework with quadratic fitting surrogates for the objective function [4,5]. The principal operation of the method relies on building, successively updating and optimizing quadratic surrogate models of the objective function over trust regions. The quadratic surrogate models reasonably reflect the local behavior of the objective function in a trust region around the current iterate and they are optimized instead of the objective function over trust regions.

Truncated conjugate gradient method by Steihaug [9] is used to find the optimal point within each trust region. The approach constructs the initial quadratic surrogate model using few data points of order $O(n)$. In each iteration of the proposed approach, the surrogate model is updated using a weighted least squares fitting. The weights are assigned to give more emphasis to points close to the current center point. The accuracy and efficiency of the proposed approach are demonstrated by applying it on a set of classical benchmark test problems and comparisons with a recent optimization technique [6] are also included.

The linear accelerators (LINACS) provide beams of high quality and high energy in which charged particles move on a linear path and are accelerated by electromagnetic fields. The modern LINAC typically consists of sections of specially designed waveguides that are excited by RF electromagnetic fields, usually in the very high frequency (VHF) range. The accelerating structures are tuned to resonance and are driven by external, high-power RF power tubes, such as klystrons. The accelerating structures must efficiently transfer the electromagnetic energy to the beam, and this is accomplished through an optimized configuration of the internal geometry, so that the structure can concentrate the electric field along the trajectory of the beam promoting maximal energy transfer, by adding nose cones to create a region of more concentrated axial electric field as shown in Fig. 1. RF cavity analysis and design brought researchers and engineers’ attention due to its extensive applications [10–17]. Applications include: medicinal purposes in radiation therapy, food sterilization and transmute nuclear fuel waste, etc. Design tools include: the computer code SUPERFISH [18], 3-D code MAFIA [19] and CST Studio Suite [20].

Design of accelerator RF cavities may include optimization of some of cavity parameters. Among the parameters characterizing the operation of the RF cavities are, the average accelerating field $E_{acc}$, peak fields to accelerating field $(E_{peak}/E_{acc})$, quality factor, and cavity shunt impedance $R_{shunt}$ [21]. The parameters considered for optimization depend on the power level fed to the cavity, which limit the average accelerating field, where the constraints on these parameters are imposed by the application. For low power level feed, optimization may focus on maximizing shunt impedance, however for high power operation, limiting the peak fields inside is of concern in order to minimize multipacting [22]. In this work we will focus on the low power fed cavities, where maximizing the shunt impedance is of main concern and will be treated through our new optimization approach.

The new proposed trust region (TR) optimization approach is capable of solving the design problems with either 2D or 3D simulators. It is expected to work as well if a 3D simulator was employed with the expense of more computational time. Most of the accelerators use body of revolution cavity structure which can be solved as 2D structure, saving the computational resources. However, the proposed approach was successfully employed in microwave filter design utilizing 3D full-wave EM solver [23].

**The new trust region approach**

The computationally expensive objective function is locally approximated around a current iterate $x_k$ by a computationally cheaper quadratic surrogate model $M(x)$ which can be placed in the form:

![Cross section of the cavity with nose cones and spherical outer walls.](image)
where \( a \in \mathbb{R} \), the vector \( b \in \mathbb{R}^n \), and the symmetric matrix \( B \in \mathbb{R}^{n \times n} \) are the unknown parameters of \( M(x) \). The total number of the model parameters is \( q = (n + 1)(n + 2)/2 \). These parameters can be evaluated by interpolating the objective function at \( q \) points.

**Initial model**

Let \( x_0 \) be the initial point that is provided by the user. Initially, assuming that \( B \) is a diagonal matrix, then the number of points required to construct the initial model is \( m = 2n + 1 \) [7]. The initial \( m \) points \( x_i, i = 1, 2, \ldots, m \), can be chosen as follows [6, 24]

\[
x_i = x_0 \quad \text{and} \quad x_{i+n} = x_0 - \Delta_i e_i, \quad i = 1, 2, \ldots, n
\]

where \( \Delta_i \) is the initial trust region radius that is provided by the user, and \( e_i \) is the \( i \)th coordinate vector in \( \mathbb{R}^n \).

The initial quadratic model \( M^{(1)}(x) \) will have the parameters \( a^{(1)} \), the vector \( b^{(1)} \), and the \( n \) diagonal elements of the model Hessian matrix \( B^{(1)} \). These parameters are computed by requiring that the initial model interpolates the objective function \( f(x) \) at the initial \( m \) points given in (2). Therefore the initial model parameters are obtained by satisfying the matching conditions:

\[
M^{(1)}(x_i) = f(x_i), \quad i = 1, 2, \ldots, m.
\]

**Model optimization**

At the \( k \)th iteration, assume that \( x_k \) is the current solution point. The model \( M^{(k)}(x) \) is then minimized, in place of the objective function, over the current trust region and a new point is produced by solving the trust region sub-problem:

\[
\min M^{(k)}(s), \quad \text{subject to} \quad ||s|| \leq \Delta_k,
\]

where \( s = x - x_k \), \( \Delta_k \) is the current trust region radius, and \( || \cdot || \) throughout is the \( l_2 \)-norm. This problem is solved by the method of truncated conjugate gradient by Steihaug [9]. It is identical as the standard conjugate gradient method as long as the iterates are inside the trust region. If the conjugate gradient method terminates at a point within the trust region, this point is a global minimizer of the objective function. If the new iterate is outside the trust region, a truncated step which is on the region boundary is considered. Also, the method treats the case where the minimum is in the opposite direction of the conjugate direction which is due to the non convexity of the model [9]. One good property of this method is that the solution computed has a sufficient reduction property, which was proved by Bandler and Abdel-Malek [25].

Let \( s^* \) denotes the solution of (4), and then a new point \( x_{k+1} = x_k + s^* \) is obtained. The achieved actual reduction in the objective function is compared to that predicted reduction using the model by computing the reduction ratio which is given by:

\[
r_k = \frac{\text{actual reduction}}{\text{predicted reduction}} = \frac{f(x_k) - f(x_{k+1})}{M^{(k)}(x_k) - M^{(k)}(x_{k+1})}.
\]

This ratio reflects how much the surrogate model agrees with the objective function within the trust region. The trust region radius and the current iterate will be updated such that, if \( r_k \) is sufficiently high, i.e., \( r_k \geq 0.7 \), there is a good agreement between the model and the objective function over this step. Hence, it is beneficial to expand the trust region for the next iteration, and to use \( x_k \) as the new center of the trust region. If \( r_k \) is positive but not close to 1, i.e., \( 0.1 \leq r_k < 0.7 \), the trust region radius is not altered. On the other hand, if \( r_k \) is smaller than a certain threshold, \( r_k < 0.1 \), the trust region radius is reduced. The updating formula used for updating \( \Delta_k \) and \( x_k \) can be expressed as follows:

\[
\begin{align*}
\text{if } r_k < 0.1: & \quad \Delta_{k+1} = \frac{1}{2} \Delta_k \\
\text{if } 0.1 \leq r_k \leq 0.7: & \quad \Delta_{k+1} = \Delta_k \\
\text{if } r_k \geq 0.7: & \quad \begin{cases}
|s^*| < \Delta_k : \Delta_{k+1} = \Delta_k \\
|s^*| \geq \Delta_k : \Delta_{k+1} = 1.5 * \Delta_k
\end{cases}
\end{align*}
\]

\[
x_{k+1} = \begin{cases}
x_k + s^* \quad \text{if } r_k > 0 \\
x_k \quad \text{otherwise}
\end{cases}
\]

It is to be mentioned that the current center is the point of least function value achieved so far.

**Model update**

When a new point is available, the current quadratic model \( M^{(k)}(x) \) is updated so that the point of lowest objective function value \( x_k \) is now the center of the \( k \)th trust region. The model will take the form:

\[
M^{(k)}(s) = d^{(k)} + s^T b^{(k)} + \frac{1}{2} s^T B^{(k)} s, \quad s = x - x_k \quad \text{and} \quad s \in \mathbb{R}^n.
\]

The parameters: \( d^{(k)} \), \( b^{(k)} \) and \( B^{(k)} \) are evaluated employing the parameter values of the previous model \( M^{(k-1)}(x) \) in addition to all available function values. The constant \( d^{(k)} \) is assigned the value of \( f(x_k) \), i.e., \( d^{(k)} = f(x_k) \). The model will be updated in two steps. First, the vector \( b^{(k)} \) is updated then the Hessian matrix \( B^{(k)} \) is updated as follows:

**Step 1:** Updating the vector \( b^{(k)} \)

The vector \( b^{(k)} \) can be obtained using only \( n \) points. However, using the \( n \) recent points may result in ill-conditioned system of linear equations. In order to avoid this, it is proposed to use the least squares approximation with the most recent \( 2n \) points. So, the vector \( b^{(k)} \) is evaluated such that the model \( M^{(k)}(x) \) fits the last \( 2n \) points obtained, \( x_i, i = 1, 2, \ldots, 2n \), i.e., the following condition should be satisfied:

\[
M^{(k)}(s_i) = f(s_i), \quad \text{where } s_i = x_i - x_k \quad \text{and} \quad i = 1, 2, \ldots, 2n.
\]

When computing the vector \( b^{(k)} \), the matrix \( B^{(k-1)} \) is assigned temporarily the value of the previous model Hessian matrix, \( B^{(k-1)} \), hence the vector \( b^{(k)} \) is obtained by solving the following system of linear equations:

\[
Ab^{(k)} = r,
\]

where
\[
A = \begin{bmatrix}
  s_1^T \\
  s_2^T \\
  \vdots \\
  s_{2n}^T
\end{bmatrix}
\quad \text{and} \quad
r = \begin{bmatrix}
  f(s_1) - a(s) - \frac{1}{2}s^T B(k-1)s_1 \\
  f(s_2) - a(s) - \frac{1}{2}s^T B(k-1)s_2 \\
  \vdots \\
  f(s_{2n}) - a(s) - \frac{1}{2}s^T B(k-1)s_{2n}
\end{bmatrix}.
\]  

(11)

The previous system is an over-determined system. The least squares approximation for \( b(k) \) is

\[
b(k) = (A^T A)^{-1} A^T r. \tag{12}
\]

**Step 2: Updating the matrix \( B(k) \)**

The model Hessian matrix \( B(k) \) is evaluated using the following updating formula:

\[
B(k) = \begin{bmatrix}
  c & \mathbf{q}^T \\
  \mathbf{q} & \mathbf{q}\mathbf{q}^T
\end{bmatrix} - \mathbf{q}\mathbf{r}, \tag{13}
\]

where \( c \) is a positive constant, \( 0.5 < c < 1 \), and the vector \( \mathbf{p} \in \mathbb{R}^n \),

\[
\mathbf{q} = \text{sign} \left( \text{diag} \left( B(k-1) \right) \right) + \sqrt{1 - c * \text{diag} \left( B(k-1) \right)}. \tag{14}
\]

This choice of \( \mathbf{q} \) ensures that changes in \( B(k) \) occur gradually. The vector \( \mathbf{p} \) is evaluated such that the model \( M^0(x) \) tries to fit all the available \( m \) points obtained so far, \( x_i, i = 1, 2, \ldots, m \), i.e., the following condition should be satisfied

\[
M^0(x_i) = f(x_i), \quad \forall x_i, \quad i = 1, 2, \ldots, m, \tag{15}
\]

i.e., the vector \( \mathbf{p} \) is obtained by solving the weighted system of linear equations

\[
A \mathbf{p} = \mathbf{r}, \tag{16}
\]

where

\[
A = \begin{bmatrix}
  \frac{1}{2}s_1^T \mathbf{q}^T \mathbf{w}_1 \\
  \frac{1}{2}s_2^T \mathbf{q}^T \mathbf{w}_2 \\
  \vdots \\
  \frac{1}{2}s_{2n}^T \mathbf{q}^T \mathbf{w}_{2n}
\end{bmatrix},
\]

\[
\mathbf{r} = \begin{bmatrix}
  w_1 \ast (f(s_1) - a(s) - \frac{1}{2}s^T B(k-1)s_1) \\
  w_2 \ast (f(s_2) - a(s) - \frac{1}{2}s^T B(k-1)s_2) \\
  \vdots \\
  w_{2n} \ast (f(s_{2n}) - a(s) - \frac{1}{2}s^T B(k-1)s_{2n})
\end{bmatrix}. \tag{17}
\]

To obtain more accurate model in the neighborhood of the current center, the available points are assigned different weights \( w_i, i = 1, 2, \ldots, m \) according to their distances from the trust region center. In the proposed approach the weight \( w_i \) associated with each equation, takes the form:

\[
w_i = \begin{cases}
  1 & \text{if } ||s_i|| \leq c_1 \Delta \\
  \frac{c_1 \Delta}{||s_i||} & \text{if } ||s_i|| > c_1 \Delta, i = 1, 2, \ldots, m.
\end{cases} \tag{18}
\]

where \( c_1 \) is a positive constant, \( c_1 \geq 1 \).

The previous system in (16) is an over-determined system \((m > n)\). The least squares approximation for \( \mathbf{p} \) is

\[
\mathbf{p} = (A^T A)^{-1} A^T \mathbf{r}. \tag{19}
\]

After getting the vector \( \mathbf{p} \), the term \( \mathbf{q}\mathbf{p}^T \) is calculated and the matrix is made symmetric by resetting the off-diagonal elements to their average values, i.e., \( b_{ij} = b_{ji} = (b_{ij} + b_{ji})/2 \), then the new Hessian matrix \( B(k) \) is updated according to Eq. (13).

The model can be improved by generating a new point \( x_{\text{new}} = x_{\text{opt}} - s_k \), which is chosen to be on the boundary of the trust region so that it improves the distribution of points around the center of the trust region. A suggested solution to find \( x_{\text{new}} \) is to solve the following problem:

\[
\max_p \sum_{i=1}^{n} \left( s_i^T s \right)^2, \quad \text{such that } s_i^T s < 0 \quad \forall i \text{and } ||s|| \leq \Delta. \tag{20}
\]

where \( s_{\text{new}} \) is selected to maximize the sum of squares of the projections of the vector \( s_{\text{new}} \) on the other \( s_i, i = 1, 2, \ldots, n \), vectors, where \( n \) is the available set of points. After generating \( s_{\text{new}} \), the function value \( f(x_{\text{new}}) \) is computed. If \( f(x_{\text{new}}) \) is found to be less than \( f(x_k) \), then \( x_{\text{new}} \) will be considered as the new trust region center of the subsequent iteration, otherwise, \( x_{\text{new}} \) will just be added to the available set of points.

### Algorithm

A complete algorithm for the proposed method is given below (see also an illustrative flowchart in Fig. 2).

1. Set \( N = 0 \) (the number of function evaluations), given \( x_0 \in \mathbb{R}^n, \Delta_1 > 0, 0.5 < c < 1, c_1 \geq 1, n_{\text{max}}, \delta \) (a termination criterion).
2. Find the initial \( m \) points using (2), letting \( x_1 \) be the initial trust region center, then construct the initial quadratic model using (3), Set \( k = 1 \).
3. Solve the trust region sub-problem (4) using the truncated conjugate gradient method to obtain \( s^* = x_k - x_0 \) of the model \( M^0(x) \) over the trust region.
4. Evaluate \( f(x_k) \) and compute the reduction ratio by substituting in (5).
5. Update the trust region radius to obtain \( \Delta_{k+1} \) using (6).
6. Determine the trust region center of the next iteration \( x_{k+1} \) based on \( x_0 \) and \( r_k \) using (7). If \( ||f(x_{k+1}) - f(x_k)|| < \delta \), the algorithm will be terminated with \( x_{\text{opt}} = x_{k+1} \) and \( f_{\text{opt}} = f(x_{k+1}) \). If for two successive iterations, \( r_k \) is negative go to Step 9, else continue.
7. Add the point \( x_k \) to the set of available points \( S \), if the number of points in \( S \) exceeds \( n_{\text{max}} \), remove the farthest point from \( x_{k-1} \).
8. Construct the quadratic model \( M^{k+1}(x) \) around \( x_{k+1} \) based on \( M^0(x) \) and the set of available points \( S \) using the updating procedures in Eqs. (9)-(19), then set \( k = k + 1 \) and go to Step 3.
9. Generate a new point \( s_{\text{new}} \) using (20), add it to the set of points \( S \), then go to Step 8.

### Examples

The effectiveness of the proposed algorithm is demonstrated through two benchmark examples. All results are compared with those obtained by NEWUOA (NEW Unconstrained Optimization Algorithm) by Powell [6]. The performance is measured by the number of function evaluations \( N \) required to reach the optimal solution.

### The 2D Beale function

The function is by [26]:

\[
f(x) = \sum_{i=1}^{3} a_i \left( x_i - x_i^2 \right)^2, \tag{21}
\]

where \( a_1 = 1.5, a_2 = 2.25, \) and \( a_3 = 2.625 \). This function has a valley approaching the line \( x_2 = 1 \), and has a minimum of 0 at \((3.05)^2\). The initial values used for \( x_0 \) and \( \Delta_1 \) are \((0.1 \ 0.1)^T\) and
0.8, respectively. The results in Table 1 and Fig. 3 compare the optimal value obtained by applying the proposed technique versus NEWUOA with the same number of function evaluation $N$.

It is to be noticed that starting from the same initial point and after only 11 iterations; the proposed algorithm gives a function value of 0.8065 while NEWUOA gives 14.2031.

The 3D Box function

The function was proposed by [27]:

$$f(x) = \sum_{i=1}^{10} \left( \exp \left( -\frac{-|x_i|}{10} \right) - \exp \left( -\frac{|x_i|}{10} \right) \right) - x_1 \left[ \exp \left( -\frac{x_1}{10} \right) - \exp(-1) \right]$$

(22)

This function has a minima at $(1, 10, 1)^T$, and also along the line{$(a, a, 0)^T$} with value 0. The initial values used for $x_0$ and

**Fig. 2** A flowchart for the proposed optimization algorithm.
Table 1 Results of the 2D Beale function compared with NEWUOA.

| N   | Proposed algorithm | NEWUOA   |
|-----|--------------------|----------|
| 11  | 0.8065             | 14.2031  |
| 21  | 0.1083             | 0.91702  |
| 31  | 0.0033             | 0.03436  |
| 43  | 2.3335e-5          | 1.7965e-4|
| 55  | 2.6973e-6          | 6.5829e-11|
| 67  | 2.5790e-7          | 6.4829e-11|

Fig. 3 Results of the 2D Beale function.

Table 2 Results of the 3D Box function compared with NEWUOA.

| N   | Proposed algorithm | NEWUOA   |
|-----|--------------------|----------|
| 10  | 0.2413             | 0.59732  |
| 17  | 5.2048e-3          | 0.18785  |
| 25  | 2.3149e-3          | 0.11451  |
| 38  | 4.2472e-4          | 0.26465e-1|
| 48  | 4.1820e-5          | 0.24613e-1|
| 62  | 4.1771e-6          | 0.21593e-2|
| 87  | 1.9203e-6          | 6.975e-5  |

Fig. 4 Results of the 3D Box function.

Updated during the optimization process, and a few iterations in the optimization process will be sufficient at the beginning. In this case the new technique will produce a significant reduction in few iterations.

Optimized design of RF cavity

The RF cavity is a major component of linear accelerators [30,31]. The structure of RF cavity must efficiently transfer the electromagnetic energy to the charged particles beam. This can be accomplished through an optimized configuration of its internal geometry, by adding nose cones to create a region of more concentrated axial electric field along the path of the electron beam, as shown in Fig. 1.

The most useful figure of merit for high field concentration along the beam axis and low ohmic power loss in the cavity walls is the effective shunt impedance per unit length $ZT$ where $T$ is the transient-time factor (a measure of the energy gain reduction caused by the sinusoidal time variation of the field in the cavity, [32]).

One of the main objectives in cavity design is to choose geometry to maximize effective shunt impedance per unit length. This indicates increasing the energy delivered to the beam compared to that thermally lost in the cavity walls. The effective shunt impedance per unit length is usually expressed in mega ohms per meter and is defined by

$$ZT = \frac{(V_0 T)^2}{PL},$$

where $P$ is the thermal power losses in the walls of the cavity, $V_0 = \int E(z)dz = E_0 L$, and $E_0$ is the average axial electric field along the cavity axis with length $L$.

The technique is applied to an RF cavity with resonance frequency 9.4 GHz, shown in Fig. 5. The objective is to maximize effective shunt impedance per unit length. In order to do that, we optimize the axial z positions of ten points that describe the cavity curvature through a spline curve. The axial positions $z = (z_1, z_2, \ldots, z_{10})^T$ in the z-direction are taken as the design parameters. The radial positions of these points are chosen on a logarithmic scale along r-direction. It is to be noted that during the variation of the curvature, the
The resonance frequency is always kept at 9.4 GHz. The initial values used for the ten radial positions $z_0$ are all set to 0.6 cm and $\Delta r$ is set to 0.02 cm.

Cavity design generally requires electromagnetic field-solver that solves Maxwell equations numerically for the specified boundary conditions. In the simulations, POISSON and SUPERFISH are used as the main solver programs in a collection of programs from LANL [18,33]. The solver is used to calculate the static magnetic and electric fields and radio-frequency electromagnetic fields for either 2-D Cartesian coordinates or axially symmetric cylindrical coordinates. The code SUPERFISH is used to solve for axisymmetric TM0nl modes, for the field components $H_{\phi}$, $E_r$, and $E_z$. The solution is obtained through solving Helmholtz equation using finite element method FEM over a triangular mesh subject to the proper boundary conditions and symmetries imposed [34].

Design algorithm shown in Fig. 6 is implemented in MATLAB code, where an initial case is chosen corresponding to ten $z$ positions of points with cavity curvature is described with spline curve (step 2). Then the spline interpolated curve is sampled at 100 points, where those sampled points are considered connected with piecewise linear, approximating the cavity curvature. This piecewise linear description is fed to AUTOMESH program to generate mesh (step 3). The solution of lowest TM mode of the cavity is made at step 4 by calling SUPERFISH, and the obtained frequency in step 5 is used to scale the cavity dimensions to keep the resonance frequency at 9.4 GHz (step 6). The corresponding scaling is reflected on the obtained cavity shunt impedance (step 7), where this value is fed to the optimizer algorithm to determine the new ten points positions. Then the process is repeated starting from step 2.

The results of the effective shunt impedance per unit length for RF cavity in mega ohm per meter after $N$ function evaluations for both the proposed algorithm and NEWUOA are shown in Table 3.

It is to be mentioned that starting from the same initial point, the convergence of the proposed algorithm is as best as NEWUOA. However, the advantage of the proposed algorithm is its easy implementation and accessibility for update and modification.

**Table 3** Results of the RF cavity design compared with NEWUOA.

| $N$  | Proposed algorithm | NEWUOA   |
|------|--------------------|----------|
| 50   | 111.771            | 112.587  |
| 75   | 115.207            | 116.833  |
| 90   | 117.183            | 119.316  |
| 120  | 119.01             | 120.511  |
| 160  | 120.5              | 120.910  |
| 200  | 121.01             | 121.211  |
| 260  | 121.301            | 121.521  |

Fig. 5 Structure of radio frequency (RF) cavity.

Fig. 6 The Poisson Superfish Solver within the proposed optimization (design) loop.

The figures of optimal cavity using the proposed algorithm and the NEWUOA are shown in Figs. 7 and 8 respectively. It worth mentioning that one could criticize the proposed optimized structure, that it contains sharp edge nose, which is difficult to manufacture and is a point of field singularity that causes breakdown. One way to override that problem is to add some curvature to the nose sharp tip, which would slightly reduce the realized shunt impedance.
Fig. 7 The optimized cavity using the proposed algorithm. Effective Shunt impedance per unit length = 121.301 MOhm/m.

Fig. 8 The optimized cavity using NEWUOA. Effective Shunt impedance per unit length = 121.521 MOhm/m.
A novel derivative-free surrogate model on the optimization of the RF cavity resonators. National Accelerator Laboratory for the fruitful discussion subjects is the major part of any linear accelerator. To obtain the optimal design for the structure of RF cavity which accuracy. In addition, the proposed method has been used to optimization or objectives that suffer from numerical inaccuracy. Thus, the proposed method is suitable for stochastic functions. Then, the inaccurate solution in case of explicit objective solution when only a limited number of function evaluations is permissible and when a high accuracy is not really necessary. Least-squares fitting is used instead of interpolation which explains the inaccurate solution in case of explicit objective functions. Thus, the proposed method is suitable for stochastic optimization or objectives that suffer from numerical inaccuracy. In addition, the proposed method has been used to obtain the optimal design for the structure of RF cavity which is the major part of any linear accelerator.

Conclusions

In this article, a new trust region optimization method that does not require any derivative information has been proposed. In this method, the objective function is approximated via quadratic surrogates, and using few number of initial data points than the exact number of surrogate parameters. Classical benchmark test problems were used to demonstrate the accuracy and efficiency of the proposed method. The results obtained showed the ability of the proposed method to rapidly converge to the final region containing the optimum solution when only a limited number of function evaluations is permissible and when a high accuracy is not really necessary. Least-squares fitting is used instead of interpolation which explains the inaccurate solution in case of explicit objective functions. Thus, the proposed method is suitable for stochastic optimization or objectives that suffer from numerical inaccuracy. In addition, the proposed method has been used to obtain the optimal design for the structure of RF cavity which is the major part of any linear accelerator.

Conflict of interest

The authors have declared no conflict of interest.

Compliance with Ethics Requirements

This article does not contain any studies with human or animal subjects.

Acknowledgment

The authors deeply thank Dr. Sami Tantawi from SLAC National Accelerator Laboratory for the fruitful discussion on the optimization of the RF cavity resonators.

References

[1] Hassan ASO, Abdel-Malek HL, Rabie AA. Non-derivative design centering algorithm using trust region optimization and variance reduction. Eng Opt 2006;38:37–51.
[2] Hassan ASO, Mohamed ASA, El-Sharabasy AY. Statistical microwave circuit optimization via a non-derivative trust region approach and space mapping surrogates. IEEE MTTS international microwave symposium digest, Baltimore, USA; June 2011. p. 1–4.
[3] Graeb H. Analog design centering and sizing. Springer; 2007.
[4] Conn AR, Toint PhL. An algorithm using quadratic interpolation for unconstrained derivative free optimization. In: Di Pillo G, Giannessi F, editors. Nonlinear optimization and applications; 1996. p. 27–47.
[5] Powell MJD. UOBYQA: unconstrained optimization by quadratic approximation. Math Program 2002;92:555–82.
[6] Powell MJD. The NEWUOA software for unconstrained optimization without derivatives. Large-scale nonlinear optimization. Springer; 2006. pp. 225–297.
[7] Powell MJD. A view of algorithms for optimization without derivatives. Technical report NA 2007/03. University of Cambridge, Department of Applied Mathematics and Theoretical Physics, Cambridge, England; 2007.
[8] Nocedal J, Wright SJ. Numerical optimization. Springer Verlag; 1999.
[9] Steihaug T. The conjugate gradient method and trust regions in large scale optimization. SIAM J Numer Anal 1983;20:626–37.
[10] Le Floch J-M, Fan Y, Aubourg M, Cros D, Carvalho NC, Shan Q, et al. Rigorous analysis of highly tunable cylindrical Transverse Magnetic mode re-entrant cavities. Rev Sci Inst 2013;84(12):125114.
[11] Barbarotti S, Grimm C, Champion M, Foley M, Ginsburg CM, Gonin I, et al. Status of the mechanical design of the 650 MHz cavities for project X. In: Proceedings of particle accelerator conference, NY, USA; 2011.
[12] Belomestnykh S, Shemelin V. High-beta cavity design—a tutorial. In: SRF2005 Workshop. Ithaca, NY, USA: Cornell University; 2005.
[13] Holla M, Trabia M, Schill Jr., RA. Optimization of a five-cell niobium cavity. The nuclear science and technology division: Transmutation Sciences Materials (TRP); 2003.
[14] Cheng C, Savage P, Kurup A, Pozimski J, Letchford A. RFQ Cold Model RF measurements and waveguide-to-coaxial line transition design for the front-end test stand at RAL. In: Proceedings of particle accelerator conference, PAC; 2007. p. 1655–57.
[15] Lau WK, Chang LH, Wang CH, Yeh MS. Analysis of the superconducting RF cavity under heavy beam loading at Taiwan light source. The 8th European particle and accelerator conference (EPAC); June 2002.
[16] Prasad M, Mishra D, Bagdual PS, Hannurkar PR. RF Design and characterization of indus-1 RF cavity. In: Proc of Indian particle accelerator conference., In PAC; 2009.
[17] Yamamoto T, Dobashi K, Natsui T, Usaka M. Design of 9.4 GHz 950 kV X-band linac for nondestructive testing. In: The 10th European conference, EPAC 2006, Edinburgh, UK; June 26-30, 2006. p. 2338–60.
[18] Halbach K, Holsinger RF. SUPERFISH-A computer program for evaluation of RF cavities with cylindrical symmetry. Part Acel 1976;7:213–22.
[19] MAFIA. Version 4 Manual (PC Release 4.200). CST, Darmstadt, Germany; 1999.
[20] CST Studio Suite 2006B. Version 6.0, CST, Darmstadt, Germany; 2006.
[21] Facco A. Tutorial on low beta cavity design. In SRF2005 Workshop, Cornell University, Ithaca, NY, USA; 2005.
[22] Vaughan JRM. Multipactor. IEEE Trans Electron Devices 1986;33(7):1172–80.
[23] Elgenaway AE. A new approach for surrogate-based trust region optimization with applications in optimal system design. M.Sc. Thesis, Faculty of Engineering, Cairo University, Egypt; 2013.
[24] Bandler JW, Abdel-Malek HL. Optimal centering, tolerancing and yield determination using multidimensional approximation. In: Proc IEEE international symposium on circuits and systems (Phoenix, Arizona); April 1977. p. 219–22.
[25] Yuan Y. On the truncated conjugate gradient method. Technical Report, ICM99-003, ICMSEC, Chinese Academy of Sciences, Beijing, China; 1999.
[26] Beale EML. On an iterative method of finding a local minimum of a function of more than one variable. Tech. Rep. No. 25, Statistical Techniques Research Group, Princeton Univ., Princeton, N.J.; 1958.
[27] Box MJ. A comparison of several current optimization methods, and the use of transformations in constrained problems. Comput J 1966(9):1-77.
[28] Hassan ASO, Abdel-Naby A. A new hybrid method for optimal circuit design using semi-definite programming. Eng Opt 2012;44(6):725–40.
[29] Abdel-Malek HL, Hassan ASO, Soliman EA, Dakroury SA. The ellipsoidal technique for design centering of microwave
circuits exploiting space-mapping interpolating surrogates. IEEE Trans Microwave Theory Tech 2006;54(10):3731–8.

[30] Nimje VT, Bhattacharjee D, Dixit K, Jayaprakash D, Mittal KC, Ray AK. Design and development of 30 MeV 3KW RF electron linac. In: Asian particle accelerator conference, APAC 2007; Indore, India; 2007. p. 491–3.

[31] Pichoff N. Introduction to RF linear accelerators. CERN 105–128, 2006.

[32] Wangler T. Principles of RF linear accelerators. John Wiley & Sons, Inc.; 1998.

[33] Los Alamos National Lab (LANL). Los Alamos Accelerator Code Group (LAACG); 2011. <laacg.lanl.gov/laacg/services/download_sf.phtml>.

[34] Menzel MT, Stokes HK. User’s guide for the poisson/superfish group of codes. NM (United States): Los Alamos National Lab.; 1987, LA-UR–87–115, January.