Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

The Regional Conjugate Optimisation Approach: A Novel Method for Three Parameter Identification of Physiological Models

Shaun M. Davidson, Paul D. Docherty and Peter A. Allison

Department of Mechanical Engineering, University of Canterbury, Christchurch, New Zealand (e-mail: shaun.davidson@pg.canterbury.ac.nz)

Abstract: Gradient descent parameter identification methods are typically effective for objective surfaces that are well conditioned, but can result in excessive computational cost or failed convergence when this is not the case. This research presents a novel Regional Conjugate Optimisation (RCO) approach, for parameter identification in three dimensions. The method characterises the objective surface prior to the iterative process, which improves the ability of the method to correctly determine key features of the objective surface and exploit these characteristics during iteration. The RCO method is validated using a Monte-Carlo methodology on a series of contrived objective surfaces, and compared to the Levenberg-Marquardt (LMQ) gradient descent method. The RCO method demonstrated faster convergence in 96% of the tested cases, demonstrating the potential of this method to result in faster convergence, higher accuracy, and lower computational cost for certain classes of problems.

1. INTRODUCTION

Gradient descent parameter identification methods, such as Levenberg-Marquardt (LMQ), rely on being able to consistently determine the direction of objective surface gradient descent in order to converge to a solution (Marquardt, 1963, Levenberg, 1944). Such methods are typically effective for objective surfaces that are well conditioned, but can result in excessive computational cost or failed convergence when this is not the case. In particular, use of gradient descent methods for models with significant parameter tradeoff effects can lead to premature termination (Docherty et al., 2013, Docherty et al., 2012c, Docherty et al., 2012a). It has been shown that in such models, the premature termination most often occurs on the major axes of the objective surface contours (that are often elliptical). On such objective surfaces, gradients approach zero far from the true minima, which can result in premature declarations of convergence distant from the optimal parameter values.

Gradient descent methods typically determine the direction of descent via evaluation of the local sensitivity of the objective surface to infinitesimal changes in the parameter set through use of a Jacobian (Levenberg, 1944, Marquardt, 1963, Bard, Davidon, Steihaug). This Jacobian is most often determined via numerical construction by evaluating the objective surface at a number of locations in the proximity of the current estimated solution. As such, where sharp contours in the objective surface exist, it is possible for numerically derived Jacobians to fail to sample the surface near the contour and thus fail to detect the optimal direction of descent. This ultimately results in excessive computational cost or convergence to parameters values that are not optimal solutions. The latter is often referred to as a local minima. However, it is more often a failure of the parameter identification gradient descent.

This research presents a novel method, the Regional Conjugate Optimisation (RCO) approach, for parameter identification in three dimensions. The method characterises the objective surface prior to the iterative process such that iterations are in the direction of major and minor error reduction. This is somewhat similar to the conjugate gradient method (Shewchuk, 1994). However, the proposed method utilizes a regional rather than local approach. The characterisation of the objective surface in a regional context allows the RCO method to orient itself along the axes of the objective surface, ensuring that it is aligned with the contours of the surface regardless of how sharp these contours may be. This improves the ability of the method to correctly determine key features of the objective field as it iterates when compared to a gradient descent method. As such, the RCO method has the potential to result in faster convergence, higher accuracy, and lower computational cost for certain classes of problems.

The method is compared to the LMQ in derived objective surfaces. The comparison is made in terms of robustness, and the number of forward simulations required for convergence.

2. METHOD

2.1 Regional Conjugate Optimisation (RCO) Parameter Identification Methodology

Broadly, the RCO is a two-stage process. The first stage characterises the objective field via characterising the objective values about the perimeter of the likely parameter set. The second stage iterates to a solution via a second order Conjugate convergence approach. This article presents the method specifically for parameter identification of models with three parameters, but there is potential to generalise the approach to n dimensions.

Copyright © 2014 IFAC

Fig. 1 exhibits several key features of the first stage of the RCO approach. The objective function value is in 3D space, and is visualised in Fig.1 by a colourmap on the surface of a sphere that encompasses the likely parameter space. The RCO approach first determines error profiles along 3 orthogonal rings (denoted here in red, green and blue) and uses these profiles to approximate the major axis of the objective function in 3-space. The error profile on a ring perpendicular to this major axis (as this is the plane where the second major and minor axes lie) is then defined (denoted in yellow), and this profile used to determine the direction of these last two axes.

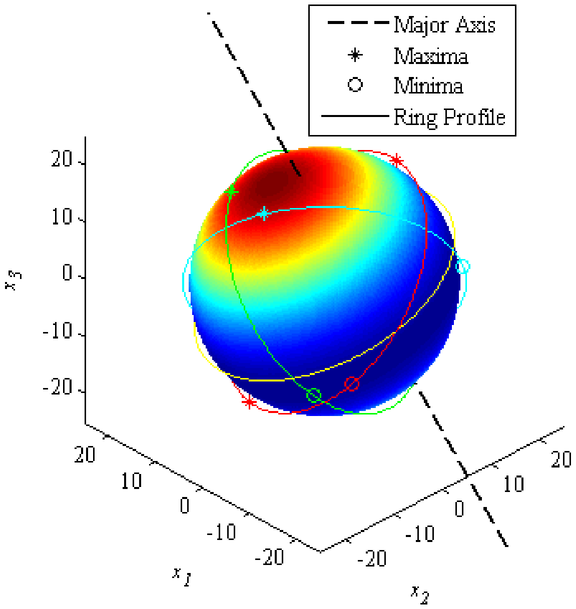

Fig. 1. Rings and Objective Surface Axes

The steps required for the RCO method to identify the optimal parameter set ($\mathbf{x} = [x_1, x_2, x_3]^T$) are as follows:

1. Define an expected parameter space x_{iE}. Scale parameters such that this space is a spherical, centred on a point x_{iC} with radius x_{iR}.

$$\mathbb{R}(x_{iE}) = [x_{i,Min}\,, x_{i,Max}] \tag{1}$$

$$x_{iC} = \frac{1}{2}\left(x_{i,Min} + x_{i,Max}\right) \tag{2}$$

$$x_{iR} = \frac{1}{2}\left(x_{i,Max} - x_{i,Min}\right) \tag{3}$$

where $i = 1, 2, 3$.

2. Define five points in the parameter space on each of three orthogonal rings C_{ij}, centred on x_{iC}, on the 1-2, 2-3 and 1-3 planes of the parameter space of radius $2x_{iR}$. Since many points occur on the intersection of two rings, this gives a total of nine points.

$$C_{12} = [2x_{1R}\cos(\omega_1) + x_{1C} \quad 2x_{2R}\sin(\omega_1) + x_{2C} \quad 0] \tag{4a}$$

$$C_{13} = [2x_{1R}\cos(\omega_1) + x_{1C} \quad 0 \quad 2x_{3R}\sin(\omega_1) + x_{3C}] \tag{4b}$$

$$C_{23} = [0 \quad 2x_{2R}\cos(\omega_1) + x_{2C} \quad 2V_{3R}\sin(\omega_1) + x_{3C}] \tag{4c}$$

where: $\omega_1 = \left[0, \frac{\pi}{4}, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\right]$

3. The value of the objective function at these points is determined $\psi(C_{i,j,\omega})$

4. A double sinusoid is used to approximate the objective values about the three rings (Equation 5, Fig. 2). Linear least squares and the $C_{i,j,\omega}$ values were used to find the value of $a_{1\text{-}5}$.

$$\psi(\omega) = a_1 sin(2\omega) + a_2 cos(2\omega) + a_3 sin(\omega) + a_4 cos(\omega) + a_5 \tag{5}$$

where: ω=[0, 2π]
and

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 2^{\frac{-1}{2}} & 2^{\frac{-1}{2}} & 1 \\ 0 & -1 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 1 \\ 0 & -1 & -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} \psi(C_{i,j,0}) \\ \psi(C_{i,j,\pi/4}) \\ \psi(C_{i,j,\pi/2}) \\ \psi(C_{i,j,\pi}) \\ \psi(C_{i,j,3\pi/2}) \end{bmatrix}$$

5. Determine the global and local objective maxima of on each ring: $\omega_{M,i,j} = \underset{\omega}{\operatorname{argmax}}\, \psi(\omega_{i,j})$

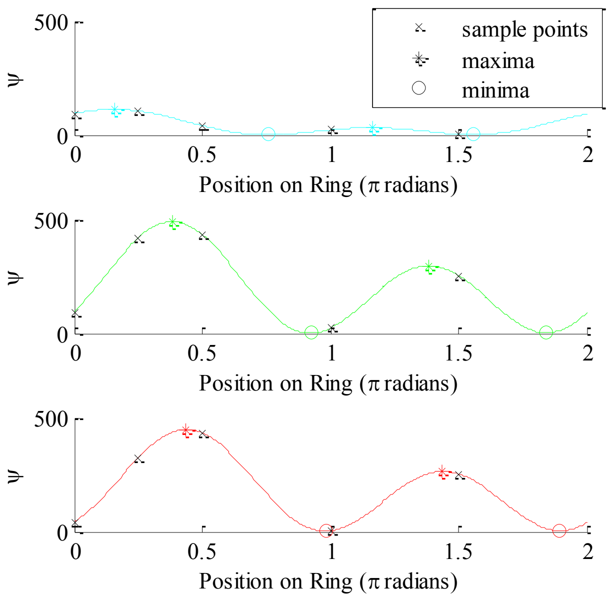

Fig. 2. Error profiles for the three orthogonal rings

6. Using Equation 6, determine the angle of the maxima-maxima line through $\omega_{M,i,j}$ for each ring $(\theta_{i,j})$.

$$\theta_{i,j} = \begin{bmatrix} \cos(\psi_{M,1,i,j}) - \cos(\psi_{M,2,i,j}) \\ \sin(\psi_{M,1,i,j}) - \sin(\psi_{M,2,i,j}) \end{bmatrix} \qquad (6)$$

   Note that at certain orientations there may only be one maxima in a profile. If there are not at least 2 profiles with both a local and global maxima (which are required to generate a maxima-maxima line) present, then one of two steps is taken. If the ratio of the sum of all maxima error values to the sum of all minima error values is greater than 5, the objective surface is deemed to be sufficiently sensitive to axis orientation. This means the objective surface is likely to have sufficiently sharp contours that it will benefit from use of the RCO approach, and that the RCO should be able to determine axis direction easily. In this case, the axes of the rings were reoriented by redefinition of the parameter axes, followed by a return to step 2. If the ratio is below 5, the surface was deemed relatively insensitive to axis orientation. This means the objective surface likely lacks sharp contours and as such convergence should be relatively independent of alignment of iterations with the axes of the objective surface, thus the code was advanced to step 10 with major, second major and minor axes assigned as [1 0 0], [0 1 0] and [0 0 1].

7. The maxima-maxima lines M_{ij} behave similarly to projections onto a plane of the major axis (Fig. 3). Any two can be combined, using the normals of the planes these lines sit on, N_{ij}, to produce an approximate 3D 'major axis' line as shown in Equation 7. Take the average direction of these 3D lines; this corresponds to the major axis direction (Fig. 2-4).

$$A_1 = (M_{ij} \times N_{ij}) \times (M_{kl} \times N_{kl}) \qquad (7)$$

   Fig. 3 shows the maxima-maxima lines generated by step 6 and the major axis line (M1) of the objective surface.

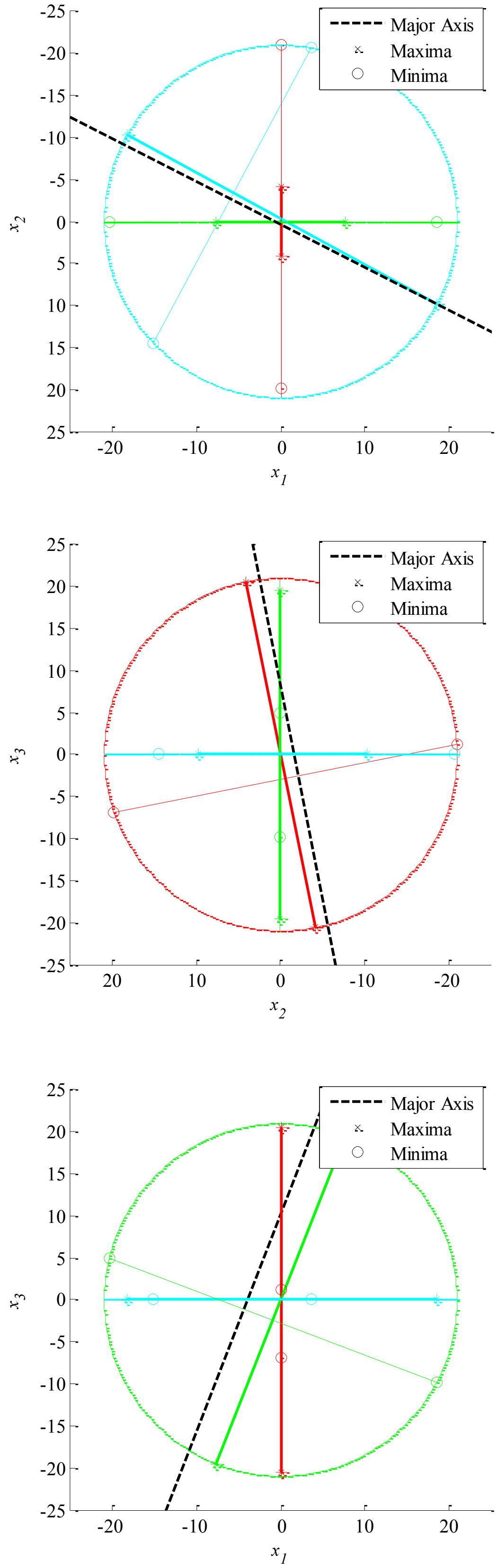

Fig. 3. Maxima-maxima and minima-minima lines for the three orthogonal rings compared to the major axis direction. The maxima-maxima (*) lines on these planes run roughly parallel to the major axis projection on these planes.

8. Determine the intersection locations of a plane that is perpendicular to this major axis and located on the centre of parameter range and the orthogonal rings. The second major and minor axes will lie on this plane. Determine the value of the objective functions derived by Equation 5 at the intersections. This should yield 6 points on a ring. Repeat steps 4-5 for this new ring (Fig. 4)

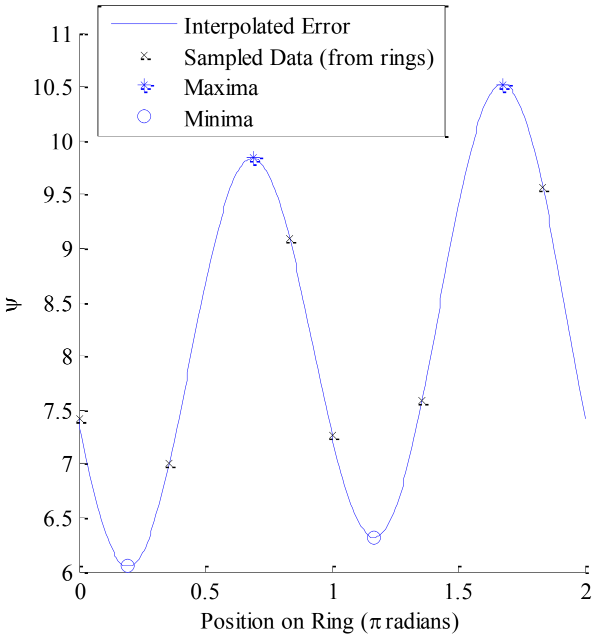

Fig. 4. Ring Error Profile Perpendicular to Major Axis

9. Take maxima-maxima and minima-minima lines from this ring. The maxima-maxima line corresponds to the second-major axis (A_2), while the minima-minima line corresponds to the minor axis (A_3).

The second stage of the RCO approach involves iteration along the three defined axis directions. This is accomplished through sampling 3 points along a particular axis, generating a second order polynomial to approximate the error profile along the axis, and locating to the minima of this polynomial. This minima in 3d—space is used as the centre-point of another 2nd order approximation on one of the orthogonal axis. This process is repeated for each axis in turn until convergence.

10. Evaluate the objective function at the parameter space centre $\mathbf{x}_C$ and the two points located on the major axis A_1 a distance of x_{iR} from $\mathbf{x}_C$. Generate a second order polynomial using these three points ($\mathbf{x}_{A1,1}$, $\mathbf{x}_{A1,C}$, $\mathbf{x}_{A1,3}$) and determine the location of its minima.

$$\begin{bmatrix} 0 & 0 & 1 \\ .25 & .5 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} \psi(\mathbf{x}_{A1,1}) \\ \psi(\mathbf{x}_{A1,C}) \\ \psi(\mathbf{x}_{A1,3}) \end{bmatrix} \quad (8)$$

$$\mathbf{x}_{A1,min} = \mathbf{x}_{A1,1} - \frac{b_2}{2b_1}\left(\mathbf{x}_{A1,1} - \mathbf{x}_{A1,3}\right) \quad (9)$$

11. Repeat the above process for the second-major A_2 and minor A_3 axes using the minima of the previous iteration ($\mathbf{x}_{A1,min}$, $\mathbf{x}_{A2,min}$) as the new centre point for sampling of the polynomial.

12. For each axis, reduce the distance x_{iR} to twice the distance between the centre of the polynomial ($A_{i,C}$) and the minima of that polynomial ($A_{i,Min}$) for the previous iteration on that axis.

13. Repeat steps 10-13 until two successive iterations are within 0.01% and convergence is declared.

2.2 *Analysis of Results*

A Monte-Carlo methodology was used to compare the convergence speed of the RCO method and the LMQ algorithm. The RCO method was developed specifically for models with high levels of parameter tradeoff. Hence, it was tested *in-silico* with contrived objective surfaces. Deviation from the parameter optima in the direction of the major axis contributed to the objective surface was 130%, 300%, or 1000% greater than an equivalent contribution due to deviation in the direction of the second axis, which in turn was 130%, 300% or 1000% of the minor axis contribution. Hence, nine ratios were possible. For each combination of axis lengths, the axes were randomly reoriented and the minima randomly located a total of 100 times. Each time both the RCO and LMQ methods were used to attain a solution. Hence, a total of 900 parameter ID cases were tested with both methods.

2.3 *In-Silico Objective Generation*

The objective surface employed was an analytically generated elliptical iso-objective surface. Similar surfaces are commonly produced by physiological models (Docherty et al.). Both the LMQ and RCO method used the same initial solution estimate, and both were considered to have failed to converge if 4000 iterations were reached. The analysis was conducted using MATLAB (R2012b 64-bit). The objective surface was determined via equations 10 and 11:

$$\mathbf{A} = [\mathbf{M_1} \quad \mathbf{M_2} \quad \mathbf{M_3}]\mathbf{D}[\mathbf{M_1} \quad \mathbf{M_2} \quad \mathbf{M_3}]^{\mathrm{T}} \quad (10)$$

$$\psi = \left(\mathbf{x} - \mathbf{x_{opt}}\right)^{\mathrm{T}} \mathbf{A}^{-1}\left(\mathbf{x} - \mathbf{x_{opt}}\right) \quad (11)$$

where $\mathbf{M}_i$ is one of the axes of the objective surface (column vector), **D** is a diagonal matrix of eigenvalues that correspond to the square of the relative axis lengths of the objective surface, **A** is a 3 x 3 matrix which characterises the shape of the objective function, $\mathbf{x}$ is the point at which the objective surface is being sampled and $\mathbf{x}_{opt}$ is the global minima point.

It is important to note that RCO methodology requires $\psi(\mathbf{x})$, but does not have direct access to $\mathbf{x}_{opt}$. This ensures applicability to real-world applications.

3. RESULTS

The RCO method successfully converged in all of the cases tested, whereas the LMQ algorithm failed to converge for 32% of the cases tested. A summary of results is presented in Table 1 below, including the median and IQR iterations to convergence for each set of axes, and the percentage of the time each method converged faster.

It is worth noting that the LMQ method had a high rate of failure in cases when large relative axis ratio lengths (1000%) were induced. This was most likely due to the LMQ being unable to correctly detect the contours of the objective surface. The vast majority of the cases in which LMQ failed to converge were in cases using dependence ratios of 1-10-100, 1-10-30 and 1-3-30. The RCO ceased to consistently outperform the LMQ method when the major axis-second major axis length ratio was 130% (the smallest ratio tested). The RCO method encountered specific difficulty in cases where the major dependence-second major dependence axis length ratios were low, but the second major-minor axis length ratio was high (1-1.3-13).

Objective Function Axes	Incidence of RCO Converging Faster than LMQ (%)	RCO Iterations Med (IQR)	LMQ Iterations Med (IQR)
1-10-100	100	36 (27-45)	DNC (DNC-DNC)
1-10-30	100	27 (24-27)	DNC (DNC-DNC)
1-10-13	100	24 (21-24)	1250 (914-1504)
1-3-30	100	45 (36-81)	DNC (DNC-DNC)
1-3-9	100	27 (27-29)	597 (526-666)
1-3-3.9	100	24 (21-27)	102 (92 -111)
1-1.3-13	93	158 (63-369)	1424 (1090-1635)
1-1.3-3.9	90	36 (29-54)	106 (98-114)
1-1.3-1.69	81	30 (27-39)	44 (40-44)

Table 1: Summary of Results of the Monte Carlo Analysis

4. DISCUSSION

Overall the RCO method demonstrated faster convergence in 96% of the cases tested, and converged in all cases (as opposed to LMQ which failed to converge in 32% of cases). The Monte Carlo analysis was designed to ensure equivalent conditions for both methodologies, thus the difference in convergence speed was the result solely of the distinct approaches used. This improvement was due to the ability of the RCO method to characterise the three-dimensional objective surface during the initial nine forward simulations it performs, and subsequently iterate along the axes of the objective surface. This allows the method to converge rapidly and stably for the class of problem being assessed. Gradient descent methods, in contrast, repeatedly analyse the objective surface local to the current parameter estimate. As such, the RCO method has a higher initial computational cost as it characterises the regional objective surface, but this rapidly pays off due to it not needing to continually reassess the surface.

The improvement in convergence produced by the RCO method becomes more pronounced the larger the relative length ratios between axes become (in particular the ratio between the major and second-major axes). This is due to the method relying on the maxima–maxima lines on two-dimensional orthogonal rings. These lines correspond more accurately to projections of the major axis onto these planes when the relative axis length ratio is larger. A low ratio can lead to ‘interference’ and thus a less accurate approximation of the direction of this axis and slower convergence relative to LMQ, as observed in the cases where the major-second major axis length ratio was 130%. However, in extremely well-conditioned objective fields with axis ratios as low as 1:1.3:1.69, the RCO method still outperformed the LMQ method in 81% of cases tested.

It is worth noting that the RCO method is typically able to provide an early indication as to how effective its performance will be via the degree of scatter in the three major axis approximations derived in step 7. If these approximations agree well, the approximate major axis is likely accurate, while if there is a large degree of scatter it is likely that the major axis orientation will be poorly estimated. This provides an opportunity after only nine forward simulations to re-orient axes or modify the expected parameter space size to increase accuracy and address this issue, reducing computational load.

It is also worth noting that in determining the major axis first, the RCO method results in compounded error when determining the minor axis. As the minor axis is the axis upon which gradients are extremely shallow, accuracy in determining this axis direction is more important than accuracy in determining the direction of other axes. This can be observed in the 1-1.3-13 axis length ratio case, where the low major-second major axis length ratio leads to interference, which decreases the accuracy of the major axis approximation. This is then compounded by extremely shallow gradients along the minor axis due to the high second major-minor axis length ratio, leading to slower convergence of the RCO. As such determination of the major axis first is not ideal. However, the alignment between maxima-maxima lines and the major axis in Fig. 2, is considerably more robust, and is thus appropriate.

The analytical approximation of parameter surfaces employed are representative only of a subset of the potential error surfaces that exist. Further, the RCO method employed approximates a second order surface with a second order function, thus the method was expected to perform well. Some further work remains to generalize the RCO method so that it is applicable to a wider variety of objective surfaces. Additionally some objective surfaces contain multiple minima or minima at locations that are difficult to estimate prior to analysis. In these cases, it is likely the RCO method would lose applicability. Thus, the RCO is not currently recommended for broad application in parameter identification. However, it should be considered for cases wherein parameter trade-off is a significant factor that causes premature convergence in gradient descent approaches and thus sub-optimal parameter estimates.

The analysis has shown the RCO method is applicable and results in rapid convergence for a certain class of problem. This demonstrates the potential of the method to produce faster convergence by characterising the objective surface of a problem. However, future work remains to generalise and validate the method so that it is applicable to a wider variety of problems. This work would likely be centred on the polynomial approximation employed in step 10 of the method, and the determination of the major axis employed in step 7. There is also longer term potential to further generalise the RCO methodology to be applicable in n-dimensions.

5. CONCLUSION

A novel regional conjugate optimisation approach, for parameter identification on three dimensions was developed. This method was designed to characterise the objective surface prior to it beginning to iterate, allowing alignment with the axes of the surface. This characterisation improves the ability of the method to correctly determine key features of the objective surface, even when the surface contained sharp contours that a gradient descent method may miss.

The RCO method was validated using a Monte-Carlo methodology on a series of contrived objective surfaces, and compared to the Levenberg-Marquardt (LMQ) gradient descent method. The RCO method converged in all 900 cases tested within 4000 iterations, whereas LMQ failed to converge in 32% of cases within 4000 iterations. The RCO method demonstrated faster convergence in 96% of the tested cases. The novel methods advantages over the LMQ were more pronounced in cases where the major-second major axis length ratio of the objective surface was large.

7. CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

Bard, Y. (1970). Comparison of Gradient Methods for the Solution of Nonlinear Parameter Estimation Problems. *SIAM Journal on Numerical Analysis,* volume 7, 157-186.

Davidon, W. (1991). Variable Metric Method for Minimization. *SIAM Journal on Optimization,* volume 1, 1-17.

Docherty, P. D., Chase, J. G. and David, T. (2012a). Characterisation of the iterative integral parameter identification method. *Medical & biological engineering & computing,* volume 50, 127-134.

Docherty, P. D., Schranz, C., Chase, J. G., Chiew, Y. S. and Moeller, K. Year. Traversing the fuzzy valley: Problems caused by reliance on default simulation and parameter identification programs for discontinuous models. *In:* Proc 8th IFAC symposium on Biological and Medical Systems, August 28-31 2012b Budapest, Hungary. 490-494.

Docherty, P. D., Schranz, C., Chase, J. G., Chiew, Y. S. and Möller, K. (2012c). Traversing the Fuzzy Valley: Problems caused by reliance on default simulation and parameter identification programs for discontinuous models. volume.

Docherty, P. D., Schranz, C., Chase, J. G., Chiew, Y. S. and Möller, K. (2013). Utility of a novel error-stepping method to improve gradient-based parameter identification by increasing the smoothness of the local objective surface: A case-study of pulmonary mechanics. *Computer methods and programs in biomedicine,* volume.

Levenberg, K. (1944). A method for the solution of certain problems in least squares. *Quarterly of applied mathematics,* volume 2, 164-168.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics,* volume 11, 431-441.

Shewchuk, J. R. 1994. An introduction to the conjugate gradient method without the agonizing pain. Carnegie Mellon University, Pittsburgh, PA.

Steihaug, T. (1983). The Conjugate Gradient Method and Trust Regions in Large Scale Optimization. *SIAM Journal on Numerical Analysis,* volume 20, 626-637.