Biomedical Signal Processing and Control 30 (2016) 162–169

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

A modified approach to objective surface generation within the Gauss-Newton parameter identification to ignore outlier data points

Rebecca A.L. Gray[a], Paul D. Docherty[a,*], Liam M. Fisk[a], Rua Murray[b]

[a] Department of Mechanical Engineering, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
[b] School of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

ARTICLE INFO

Article history:
Received 30 July 2015
Received in revised form
27 December 2015
Accepted 22 June 2016
Available online 9 August 2016

Keywords:
Parameter identification
Glycemic modelling
Gradient descent methods
Gauss-Newton methods
Outlier data
Three sigma

ABSTRACT

The Gauss-Newton method is a simple iterative gradient descent method used to modify a mathematical model by minimising the least-squares residuals between the modelled response, and some observed behaviour. A common issue for parameter identification methods that optimise least-square residuals is the sporadic occurrence of outlying data in the observation data set.

This research proposes an amendment to the Gauss-Newton parameter identification approach that limits the influence of outlying data by dissipating the contribution of outlying data to the objective function that drives iterations. The modified method was tested in two and three-dimensional parameter identification exercises using virtual data from the dynamic insulin sensitivity and secretion test (DISST). The data incorporated random normally distributed noise (CV = 3%) or random normally distributed noise in concert with an outlying data point. The proposed method performed similarly to the original method when no outlying data was included and found the model that fit accurately to the majority of data points when an outlying data point was present.

The proposed approach provides a valuable tool for the rejection of outlier data that is operator independent, does not require multiple stages of analysis, or manual removal of data.

1. Introduction

Inverse problem methodologies can be used to quantify and describe physiological phenomena by fitting specific candidate models to observed behaviour [1,2]. The models are modified to capture a particular subject's observed response via the appropriate identification of key model parameter values. In particular, parameter values are identified using methods that minimise the difference between the responses simulated by the model and observed behaviour [3]. Gradient descent methods are a family of parameter identification methods that optimise the model parameters with respect to the model structure and the observed data. These methods calculate the gradient in the model-observation disagreement local to a particular parameter estimate, and iterate in a direction of reducing error. The Gauss-Newton method is a relatively simple gradient descent method from which many advanced methods are derived [4].

The Gauss-Newton method is an iterative, non-linear gradient descent method for parameter identification in terms of least squares error. This method performs well for convex systems with relatively identifiable models [5] and low levels of measurement error. The Gauss Newton method forms the foundation of the Levenberg-Marquardt algorithm [4,6,7], the Davidon-Fletcher-Powell method [4,8], and the conjugate gradient approach [4,9].

An intrinsic problem parameter identification methods that minimise squared residual error is the effect of outlier data measurements on the objective surface [10]. In such cases, outlier data can cause excessively large deviation from the behaviour described by the majority of the data points. A frequently used methodology is *a-posteriori* removal of data points that are greater than 3 standard deviations from the simulated response [11,12]. However, this method is difficult to implement, particularly in cases wherein data sets that contain sparse high accuracy measurements. The post-process method implementation also leaves scope for operator influence on the outcomes. Methods with reduced operator influence include Huber, Tukey and Cauchy adaptations to the objective contributions [13,14].

This research presents an adaption of the Gauss Newton Method that allows outlying measurements to be ignored within the identification method. Thus, statistical identification of outliers is not an aspect of post-process and thus the process is operator independent and quickly done. The adapted method has been tested in a

* Corresponding author.
E-mail address: paul.docherty@canterbury.ac.nz (P.D. Docherty).

http://dx.doi.org/10.1016/j.bspc.2016.06.009
1746-8094/

well-conditioned two dimensional case, and a poorly-conditioned three dimensional case. The method was compared to the original method for each.

2. Methods

The adapted Gauss Newton method was tested on *in-silico* data that represented a typical insulinaemic/glycemic response of a healthy participant of the dynamic insulin sensitivity and secretion test (DISST) [15–17].

2.1. The adapted Gauss-Newton parameter identification method

Gradient descent methods step iteratively towards the minimum by evaluating the direction of descending gradient. These methods typically choose a step size in proportion to the magnitude of the error. The algorithm for the Gauss-Newton method converges from a particular initial parameter estimate by iterating Eq. (1):

$$\mathbf{x}_{i+1} = \mathbf{x}_i - (\mathbf{J}_i^{\mathrm{T}}\mathbf{J}_i)^{-1}\mathbf{J}_i^{\mathrm{T}}\boldsymbol{\psi}(\mathbf{x}_i) \tag{1}$$

where $\mathbf{x}$ is the identified parameter set, i is the iteration number, $\boldsymbol{\psi}$ is the vector of residual values, and $(\mathbf{J}_i)$ is the Jacobian matrix. The values of $\mathbf{J}_i$ and $\boldsymbol{\psi}$ are defined:

$$\mathbf{J}_i = \left[\frac{\partial \psi_k}{\partial x_j}\right]_i = \begin{bmatrix} \frac{\partial \psi_1}{\partial x_1} & \frac{\partial \psi_1}{\partial x_2} & \cdots & \frac{\partial \psi_1}{\partial x_m} \\ \frac{\partial \psi_2}{\partial x_1} & \frac{\partial \psi_2}{\partial x_2} & \cdots & \frac{\partial \psi_2}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \psi_n}{\partial x_1} & \frac{\partial \psi_n}{\partial x_2} & \cdots & \frac{\partial \psi_n}{\partial x_m} \end{bmatrix}_i \tag{1a}$$

$$\psi_k(\mathbf{x}_i) = G(\mathbf{x}_i, t_k) - G_k \tag{1b}$$

where there are $j = 1 \ldots m$ model parameters to be optimised and k=1…n data measurements, $G(\mathbf{x}_i, t_k)$ are exact model solutions for G at time t_k with parameters $\mathbf{x}_i$, G_K is the observed data at times $t = t_k$ are $G_k := G_{t=t_k}$. This method ultimately leads to an optimisation in terms of least squares, and produces the objective surfaces shown in Fig. 3:

$$\mathbf{x}_{opt} = \underset{\mathbf{x}}{\operatorname{argmin}} \sum_k \psi_k^2 \tag{2}$$

However, a single outlier can have a disproportionate effect on the identified optimal parameter set. The alternative objective surface formulation used in this analysis is intended to ignore the contributions of outlier data. Fig. 1 and Eq. (3) show the original least squares objective contribution and the contrasting shape used in this analysis. The modified objective is given by:

$$\hat{\boldsymbol{\psi}} = \left[\hat{\psi}_k\right] = \psi_k(\mathbf{x}) e^{-\alpha|\psi_k(\mathbf{x})|} \tag{3}$$

The value of α can be adapted during the parameter identification iterations to reduce the width of the peak ring about the minima. This allows a wider ring at the beginning of the iterative process and a tighter ring once the method begins to converge. The median residual error $\left(|\psi|_M\right)$ typically reduces during convergence, and α is specified by Eq. (4) (with an additional tunable parameter β):

$$\alpha = \frac{1}{\beta|\boldsymbol{\psi}|_M}. \tag{4}$$

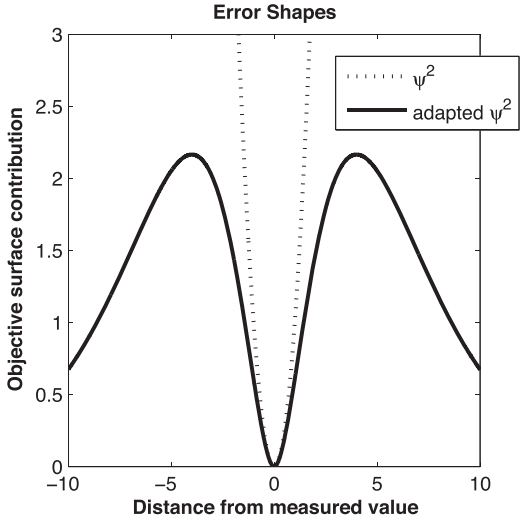

Fig. 1. The objective surface contributions from Eq. (3) with $\alpha = 0.3$.

Incorporating Eqs. (3) and (4) into Eq. (1) thus adapts the Gauss-Newton algorithm:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - (\mathbf{J}_i^{\mathrm{T}}\mathbf{J}_i)^{-1}\mathbf{J}_i^{\mathrm{T}}\hat{\boldsymbol{\psi}} \tag{5}$$

$$\text{where : } \hat{\boldsymbol{\psi}} = \left[\hat{\psi}_k\right] = \psi_k(\mathbf{x}_i) e^{-\frac{|\psi_k(\mathbf{x}_i)|}{\beta|\boldsymbol{\psi}(\mathbf{x}_i)|M}}. \tag{5a}$$

The typical $\boldsymbol{\psi}$ determined in Eq. (4) is used to find the Jacobian as this determines the direction of the next step of the Gauss-Newton method (i.e. $\mathbf{J} = f(\boldsymbol{\psi}), \mathbf{J} \neq f(\hat{\boldsymbol{\psi}})$), this is in contrast to similar methods that address outlier data such as Cauchy, Huber, Tukey [13,14]). The modulated $\hat{\boldsymbol{\psi}}$ is only used in the adapted Gauss-Newton algorithm to determine how each residual error value contributes to the magnitude of the step. This modified Gauss-Newton iteration inherits the robustness properties and mathematical proof of the typical Gauss-Newton least-squares, provided that the number of inlier data points, meets or exceeds the number of identified parameters. This is a standard requirement of Gauss-Newton [18], and mathematical proof of the typical Gauss-Newton method [18]. In particular, if the proposed method determines that the k^{th} data point is an outlier, it will reduce, or effectively eliminate the k^{th} element of the $\hat{\boldsymbol{\psi}}$ vector and will thus ameliorate the contribution of the k^{th} column of $\left[(\mathbf{J}_i^{\mathrm{T}}\mathbf{J}_i)^{-1}\mathbf{J}_i^{\mathrm{T}}\right]$

A range of values for β were considered. It was determined via inspection that the value of β was effectively equal to the number of standard deviations of the error distribution between the peaks of the error shape. An accepted statistical approach with regards to the rejection of outlier data is to manually remove data that falls more than three standard deviations away from the identified model simulation. Hence, to ensure that data within three standard deviations is captured by the adapted approach, a value of $\beta = 3$ was used. This meant that assuming normal distribution, 99.7% of all data points had significant contributions to the identified optima. Fig. 2 shows the effect of β on the width of the objective contribution function, and the effect of the measurement precision on the objective contribution function.

2.2. Construction and interpretation of the objective surface

The DISST test is a low dose, moderately sparsely sampled test for insulin sensitivity (S_I). The fasted participant receives a 10 g glucose bolus at $t = 6$ min and a 1U insulin bolus at $t = 16$ min. In this application, glucose, insulin and C-peptide samples were taken at