Mathematical Biosciences 285 (2017) 119–127

Contents lists available at ScienceDirect

Mathematical Biosciences

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

The dimensional reduction method for identification of parameters that trade-off due to similar model roles

Shaun M. Davidson[a], Paul D. Docherty[a], Rua Murray[b]

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

ARTICLE INFO

Article history:
Received 24 July 2016
Revised 25 November 2016
Accepted 13 January 2017
Available online 18 January 2017

Keywords:
Inverse problems
System identification
Parameter trade-off
Hyperplane
Gradient descent

ABSTRACT

Parameter identification is an important and widely used process across the field of biomedical engineering. However, it is susceptible to a number of potential difficulties, such as parameter trade-off, causing premature convergence at non-optimal parameter values. The proposed Dimensional Reduction Method (DRM) addresses this issue by iteratively reducing the dimension of hyperplanes where trade off occurs, and running subsequent identification processes within these hyperplanes. The DRM was validated using clinical data to optimize 4 parameters of the widely used Bergman Minimal Model of glucose and insulin kinetics, as well as *in-silico* data to optimize 5 parameters of the Pulmonary Recruitment (PR) Model. Results were compared with the popular Levenberg–Marquardt (LMQ) Algorithm using a Monte-Carlo methodology, with both methods afforded equivalent computational resources. The DRM converged to a lower or equal residual value in all tests run using the Bergman Minimal Model and actual patient data. For the PR model, the DRM attained significantly lower overall median parameter error values and lower residuals in the vast majority of tests. This shows the DRM has potential to provide better resolution of optimum parameter values for the variety of biomedical models in which significant levels of parameter trade-off occur.

1. Introduction

Parameter identification is the process of identifying the set of parameters for a model that optimize a cost or error function. Cost functions are more commonly employed during optimisation of control systems, [11]. Error function values (ψ) are minimized to find model parameters that best align a model with measured data, The resulting parameters can be used either for diagnostic purposes [10,20] or for predictive modelling or extrapolation [6,17]. These two broad cases encompass a wide variety of valuable biomedical applications, across a number of biological systems, including the circulatory [17], glycemic [10,11] and pulmonary [20] systems.

A core parameter identification method is iterative gradient descent. Here, the gradient is determined via a Jacobian ($d\boldsymbol{\psi}/d\mathbf{y}$) evaluated at a certain point in the parameter space ($\mathbf{y}$) to determine the direction that gives the steepest decrease in ψ [2,21]. The method then advances in the parameter space in the direction of this gradient before re-evaluating the Jacobian for the next iteration. Gradient descent methods are considered to be relatively robust, but can converge extremely slowly if positioned far from the optimum point [15]. When iterative gradient descent is coupled with a weighted Gauss-Newton algorithm, gradient descent becomes the popular Levenberg–Marquardt (LMQ) algorithm [14,15], which has attained a pseudo gold-standard status in parameter identification.

Parameter identification can suffer from a number of defects which result in premature declarations of convergence and failure to correctly identify the optimum parameter values. Such failures can occur due to parameter trade-off, where two or more parameters can define the same characteristics in observed behaviour [1,5,7,9] (i.e. $d\boldsymbol{\psi}/dy_1 \approx d\boldsymbol{\psi}/dy_2$). A small degree of this behaviour occurs in most models and only becomes deleterious when the trade-off is strong. Significant parameter trade-off can lead to changes in a parameter being offset by changes in another while maintaining a similar ψ minima, resulting in elongated iso-error contours and regions of very low error gradients. This phenomenon can cause parameter identification methods to determine that any point within this region of extreme parameter trade-off is an optima due to difficulties in detecting relatively low error gradients [7,8].

E-mail addresses: shaun.davidson@pg.canterbury.ac.nz (S.M. Davidson), paul.docherty@canterbury.ac.nz (P.D. Docherty), rua.murray@canterbury.ac.nz (R. Murray).

http://dx.doi.org/10.1016/j.mbs.2017.01.003
0025-5564/

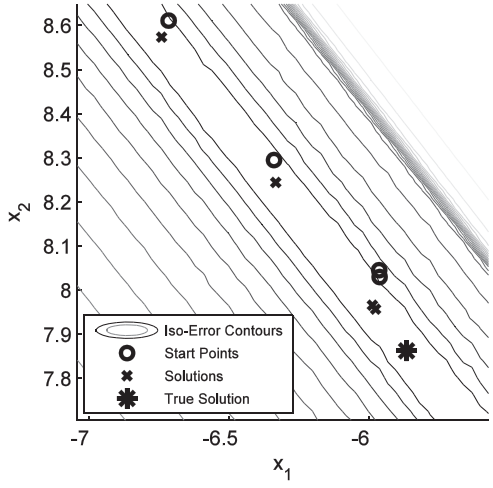

Fig. 1. Parameter trade-off in the pulmonary recruitment (PR) model.

Fig. 1 shows parameter trade-off and the associated failure to correctly detect optima during LMQ identification from near ideal starting points for the Pulmonary Recruitment (PR) Model, as observed in [8]. Despite the relatively close proximity of the start points to the true solution, the algorithm proceeds along the steep error gradients into the low error trade-off region. LMQ declares convergence once reaching this region without iterating further. This phenomenon occurs with relative frequency in parameter identification, and can sometimes be mistaken for premature convergence due to local minima. Note that these solutions lie approximately along a line of minimal error that intersects with the true model solution.

The proposed Dimensional Reduction Method (DRM) leverages the tendency of parameter trade-off to manifest on a unique $n-1$ dimensional hyperplane for an n dimensional problem (e.g. a 2D plane for a 3D problem, or a 1D line for a 2D problem, as in Fig. 1) [7]. For an n dimensional problem, the DRM method defines this hyperplane using n Levenberg-Marquardt solutions. $n-1$ Levenberg-Marquardt solutions are then found on this $n-1$ dimensional hyperplane, and an $n-2$ dimensional hyperplane is defined. This process is repeated until Levenberg–Marquardt is ultimately run in 1 dimension to determine the optimum point. By constraining gradient descent to operate only within the region of parameter trade-off, relatively large error gradients orthogonal to the hyperplane are ignored allowing easier detection of the shallower contours within the parameter trade-off region. This method is easily generalizable for any number of parameters, and has potential to be utilized as a framework for most parameter identification methods.

The DRM method is validated via a Monte-Carlo methodology across two models. The first of these models is the widely used Bergman Minimal Model of glucose and insulin kinetics [4]. Four model parameters are optimized on 36 sets of experimentally derived data across a cohort of 12 subjects. This model was selected due to it being widely recognized and employed, and the availability of experimental data allowing for validation of the DRM approach on a data set with model-data mismatch, noise, sparse data points and various other real world constraints in play.

The second model employed is the Pulmonary Recruitment (PR) model, which is a lung model that contains a large number of discontinuities due to the presence of multiple Heaviside functions and thus is difficult to accurately identify [18,19]. Five parameters of this model are optimized using *in-silico* data designed to ensure that all elements of the model are in effect. The use of *in-silico* data provides a numerically rigorous comparison between the two methods without real world constraints such as model data mismatch or unclear true parameter values. Across both of these models, the DRM method is compared to the widely used LMQ. This validation thus encompasses real data as well as simulated data and two different models with different parameter numbers ($n=4$ and $n=5$).

2. Methods

2.1. The dimensional reduction method

Note that all vectors are column vectors. A model is defined to be dependent on a vector of model parameters $\mathbf{y} \in \mathbb{R}^n$. Let $\mathbf{u}(t, \mathbf{y})$ denote the value of a measurable output $\mathbf{u}$ at time t, given parameters $\mathbf{y}$. For a set of time samples $\mathbf{t}_S = \{t_1, t_2, \ldots t_S\}$, each model solution defines a vector $\mathbf{u}(\mathbf{t}_S, \mathbf{y}) \in \mathbb{R}^S$. When a sequence of observations $\mathbf{u}_S$ at times $\mathbf{t}_S$ is available, the model error ($\boldsymbol{\psi}$) can be expressed as:

$$\boldsymbol{\psi}(\mathbf{y}) = \mathbf{u}(\mathbf{t}_S, \mathbf{y}) - \mathbf{u}_S \tag{1}$$

The parameter identification is achieved by finding

$$\mathbf{y}^{\text{opt}} = \text{argmin}_{\mathbf{y}} \left\| \boldsymbol{\psi}(\mathbf{y}) \right\|_2 \tag{2}$$

2.1.1. Initialization of the method

Assume that the vector $\mathbf{y} \in \mathbb{R}^n$ of model parameters is known to lie in a rectangular region $[y_{\min,1}, y_{\max,1}] \times \cdots \times [y_{\min,n}, y_{\max,n}]$. First transform the domain to the unit cube in $\mathbb{R}^n$ by setting

$$\mathbf{A}_0 := \text{diag}(\mathbf{y}_{\max} - \mathbf{y}_{\min}) \tag{3a}$$

$$\zeta_0 := \mathbf{y}_{\min} \tag{3b}$$

Choose n sets of random initial points $\mathbf{x}_j^{(0)} \in [0,1]^n$ $(j = 1, \ldots, n)$, and use gradient descent to compute $\text{argmin}_{\mathbf{x}} \|\boldsymbol{\psi}(\mathbf{A}_0\mathbf{x} + \zeta_0)\|_2$ (equivalent to $\|\boldsymbol{\psi}(\mathbf{y})\|_2$) from these initial guesses. Let the corresponding family of approximate optima be denoted $\{\mathbf{x}_j^{(0),\text{opt}}\}_{j=1}^n$. We assume that the affine span of these n points is an $(n-1)$-dimensional hyperplane[1,2]. These optima will enable reduction of the hyperplane dimension and continued searching.

2.1.2. First dimension reduction

To achieve the dimension reduction it is necessary to parametrize the hyperplane. Calculate a normal vector $\theta^{(n)} \in \mathbb{R}^n$ to the affine span of $\{\mathbf{x}_j^{(0),\text{opt}}\}_{j=1}^n$ such that

$$\left(\mathbf{x}_j^{(0),\text{opt}}\right)^T \theta^{(n)} = 1 \tag{4}$$

for each $j = 1, \ldots, n$. Eq. (4) provides a system of linear algebraic equations that uniquely defines $\theta^{(n)}$. Let $i^* = \text{argmax}_i |\theta_i^{(n)}|$. i^* indexes the coordinate direction in parameter space that is closest to perpendicular to the hyperplane defined by $\{\mathbf{x}_j^{(0),\text{opt}}\}_{j=1}^n$, which is the direction that corresponds to the parameter likely to be most accurately identified at this point. The hyperplane is parametrized by the $(n-1)$ coordinates excluding i^* as follows:

Let $\mathbf{B}^{(n)}$ be the $(n-1) \times n$ matrix obtained by removing the i^* row of the $n \times n$ identity matrix; that is,

$$\mathbf{B}_{ij}^{(n)} = \begin{cases} 1 & i < i^*, j = i, \\ 1 & i \geq i^*, j = i+1, \\ 0 & \text{otherwise.} \end{cases}$$

[1] The affine span of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is the set of all $\sum_{i=1}^n a_i \mathbf{x}_i$ where $\sum_{i=1}^n a_i = 1$ and each $a_i \in \mathbb{R}$.

[2] In the untypical case that the affine span has dimension less than $(n-1)$, redundant points can be removed, and additional ones generated from further random initial conditions.