The Pure and Applied Mathematics 2 (1995), No 2, pp. 133–140
J. Korea Soc. of Math. Edu. (Series B)

QR DECOMPOSITION IN NONLINEAR EXPERIMENTAL DESIGN

IM GEOL OH

ABSTRACT. The D–optimal design criterion for precise parameter estimation in nonlinear regression analysis is called the determinant criterion because the determinant of a matrix is to be maximized. In this thesis, we derive the gradient and the Hessian of the determinant criterion, and apply a QR decomposition for their efficient computations. We also propose an approximate form of the Hessian matrix which can be calculated from the first derivative of a model function with respect to the design variables. These equations can be used in a Gauss–Newton type iteration procedure.

1. Introduction

Many problems arising in science and engineering can be described by the nonlinear regression model

$$\mathbf{y} = f(\mathbf{x},\ \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$$

where $f(\mathbf{x},\boldsymbol{\theta})$ is the response function and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)^T$ is an $n \times 1$ vector of unobservable experimental errors. The response function $f(\mathbf{x},\boldsymbol{\theta})$ is a function of $\mathbf{x} = (x_1, x_2, \cdots, x_k)^T$, k–dimensional vector of independent variables, and $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_p)^T$ is a p–dimensional vector of unknown parameters.

When observed values of $\mathbf{x}$ and $\mathbf{y}$ are available, the main goal in nonlinear regression analysis is to obtain precise estimates of the model parameters $\boldsymbol{\theta}$ which can

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

be used further in predicting the future values of the response variable or in explaining the proposed functional structure. From the point of view of experimental design, however, the objective is to find "good" design points $\mathbf{x}$ under some criteria at which values of $\mathbf{y}$ are to be observed.

Several design criteria in the nonlinear regression case were proposed for each different specific purpose. For example, see Atkinson and Fedorov (1975), Box(1971), Draper and Hunter (1967), and Hill et al. (1968). For precise parameter estimation the D–optimal design criterion is proposed by Box and Lucas (1959). Suppose we want N design points $\mathbf{x}_n$, $n = 1, 2, \cdots, N$, then under this criterion, these points are chosen to minimize the determinant of the variance–covariance matrix of the parameter estimates (apart from σ^2), or, equivalently, to maximize the determinant

$$|\mathbf{V}^T\mathbf{V}| \tag{1.1}$$

where $\mathbf{V}$ is the $N \times P$ derivative matrix whose (n,p)th element is $\{\mathbf{V}\}_{np} = \frac{\partial f(\mathbf{x}_n, \boldsymbol{\theta})}{\partial \theta_p}$. This D–optimal design criterion can be interpreted as to minimize the generalized variance of the parameter estimates, or geometrically it corresponds to minimizing the volume of the joint confidence region of the parameter estimates.

Unlike the D–optimal design criterion for the linear regression models, the determinant criterion (1.1) is the function of the unknown parameters $\boldsymbol{\theta}$. Hence selection of the 'best' N design points depends, paradoxically, on the actual values of the P unknown parameters, as pointed out by Cochran (1973), Box and Lucas (1959). In practice, preliminary estimates, or estimates from the previous data, would have to be used.

Obtaining design points using the determinant criterion is a nonlinear optimization problem, in which we find the NK values of $\mathbf{x}_n$, $n = 1, 2, \cdots, N$ which maximize their nonlinear objective function $|\mathbf{V}^T\mathbf{V}|$. In order to utilize a Gauss–Newton type iteration procedure, therefore, it is necessary to develop the gradient and the Hessian of $|\mathbf{V}^T\mathbf{V}|$ with respect to $\mathbf{x}_n$. In the next section, we derive those and we suggest a compact computing scheme in Section 3. In section 4, we are reported the Concluding Remarks

2. Derivation of the Gradient and the Hessian

Collecting the N design points $\mathbf{x}_n$, $n = 1, 2, \cdots, N$, we can form an $N \times K$ design matrix $\mathbf{X} = \{x_{nk}\}$, and the determinant $|\mathbf{V}^T\mathbf{V}|$ is considered as the function of $\mathbf{X}$ with known $\boldsymbol{\theta}$. We derive the gradient and the Hessian of $|\mathbf{V}^T\mathbf{V}|$ with respect to $\mathbf{X}$ by generalizing the approach used in Bates and Watts (1985) for optimizing the multiresponse parameter estimation criterion.

2.1 Gradients.

Consider the function

$$g(\mathbf{X}) = \frac{1}{2} \ln |\mathbf{V}^T\mathbf{V}| \tag{2.1}$$

for which derivatives with respect to elements of $\mathbf{X}$ are given by (Fedorov, 1972)

$$\begin{aligned} \frac{\partial g(\mathbf{X})}{\partial x_{nk}} &= \operatorname{tr}\left((\mathbf{V}^T\mathbf{V})^{-1}\,\mathbf{V}^T\,\frac{\partial \mathbf{V}}{\partial x_{nk}}\right) \\ &= \operatorname{tr}\left(\mathbf{V}^{+}\frac{\partial \mathbf{V}}{\partial x_{nk}}\right) \end{aligned} \tag{2.2}$$

where $\mathbf{V}^{+} = (\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T$ is the pseudo-inverse of $\mathbf{V}$. Hence we obtain

$$\frac{\partial |\mathbf{V}^T\mathbf{V}|}{\partial x_{nk}} = 2\,|\mathbf{V}^T\mathbf{V}|\,\operatorname{tr}\left(\mathbf{V}^{+}\frac{\partial \mathbf{V}}{\partial x_{nk}}\right) \tag{2.3}$$

2.2 Hessian.

Four–dimensional array is needed to express the Hessian of $|\mathbf{V}^T\mathbf{V}|$ with respect to the design matrix $\mathbf{X}$. To simplify notation, we use the vec operator (Searle,

1982) which transforms a matrix to a vector by packing each column of the matrix into a vector. For example, $\mathbf{q} = \text{vec}(\mathbf{X})$ is the NK–dimensional ector whose first N elements are the first column of $\mathbf{X}$, the next N elements are the second column, and so on.

Denoting by q_t and q_s the tth and sth element of $\mathbf{q}$, the (t,s)th element of the Hessian matrix of $g(\mathbf{X})$ with respect to $\mathbf{q}$ is given by

$$\begin{aligned} g_{t,s} &= \frac{\partial^2 g(\mathbf{X})}{\partial q_t \partial q_s} \\ &= -\,\text{tr}\left[\mathbf{V}^+ \frac{\partial \mathbf{V}}{\partial q_t}\,\mathbf{V}^+ \frac{\partial \mathbf{V}}{\partial q_s}\right] \\ &+\ \text{tr}\left[\mathbf{V}^+ \,(\mathbf{V}^+)^T \left(\frac{\partial \mathbf{V}}{\partial q_t}\right)^T \left(\mathbf{I} - \mathbf{V}\mathbf{V}^T\right)\left(\frac{\partial \mathbf{V}}{\partial q_s}\right)\right] \\ &+\ \text{tr}\left[\mathbf{V}^+ \frac{\partial^2 \mathbf{V}}{\partial_t \partial q_s}\right] \end{aligned} \tag{2.4}$$

This is obtained by using the expression for the derivative of the pseudo-inverse from Golub and Pereyra (1973).

Since x_{nk} is an element of $\mathbf{q}$, we can express equation (2.2) and (2.3) as

$$\begin{aligned} g_t &= \frac{\partial g(\mathbf{X})}{\partial q_t} = \text{tr}\left(\mathbf{V}^+ \frac{\partial \mathbf{V}}{\partial q_t}\right) \\ \frac{\partial |\mathbf{V}^T\mathbf{V}|}{\partial q_t} &= 2|\mathbf{V}^T\mathbf{V}|\; g_t \end{aligned} \tag{2.5}$$

The Hessian of $|\mathbf{V}^T\mathbf{V}|$ is calculated using the relationship

$$\frac{\partial^2 |\mathbf{V}^T\mathbf{V}|}{\partial q_t \partial q_s} = 2\,|\mathbf{V}^T\mathbf{V}|\left(g_{t,s} + 2\,g_t\;g_s\right) \tag{2.6}$$

Evaluation of equation (2.4) requires the second order derivatives of the model function $f(\mathbf{x},\boldsymbol{\theta})$ with respect to the design points. Since the last term in the right-hand side of (2.4) has little effect on the overall performance of the iteration procedure, we replace the terms $\frac{\partial^2 |\mathbf{V}^T\mathbf{V}|}{\partial q_t \partial q_s}$ by zero to provide an approximate Hessian as suggested by Bates and Watts (1985). Evaluation of equations (2.5) and (2.6) can then be done by using $\mathbf{V}$ and its first derivative to the design points.

3. A Computing Scheme Using QR Decompsition

Computations of the gradient and the Hessian are simplified by considering the QR decomposition of $\mathbf{V}$ as $\mathbf{V} = \mathbf{QR}$, where $\mathbf{Q}$ is the $N \times N$ and orthogonal and $\mathbf{R}$ is $N \times P$ and zero below the main diagonal. Let $\mathbf{R}_1$ be the $P \times P$ triangular matrix, composed of the first P rows of $\mathbf{R}$ and $\mathbf{Q}_1$ the first P columns of $\mathbf{Q}$. Then $\mathbf{V}$ can be written as $\mathbf{V} = \mathbf{Q}_1\mathbf{R}_1$ and the determinant criterion is computed as

$$|\mathbf{V}^T\mathbf{V}| = |\mathbf{R}_1^T\mathbf{R}_1| = \prod_{i=1}^{P} \{\mathbf{R}_1\}_{ii}^2.$$

That is, the objective function is evaluated simply by computing the product of the square of the diagonal elements of $P \times P$ matrix $\mathbf{R}_1$, which is of very small size in most cases.

QR decomposition of $\mathbf{V}$ can also be used in computing the gradient and Hessian. First, pseudo-inverse of $\mathbf{V}$ is expressed as

$$\mathbf{V}^+ = (\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T = \mathbf{R}_1^{-1}\mathbf{Q}_1^T.$$

Gradient in equation (2.5) is then computed by, using the relationship $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$,

$$\begin{aligned}\frac{\partial|\mathbf{V}^T\mathbf{V}|}{\partial q_t} &= 2\,|\mathbf{V}^T\mathbf{V}|\ \mathrm{tr}\ (\mathbf{Q}_1^T\mathbf{V}_{(t)}\mathbf{R}_1^{-1}) \\ &= 2|\mathbf{V}^T\mathbf{V}|\sum_{i=1}^{P} c_{t,ii}\end{aligned}$$

where $\mathbf{V}_{(t)} = \frac{\partial \mathbf{V}}{\partial q_t}$ and $c_{t,ii}$ is the (i,i)th element of the $N \times P$ matrix $\mathbf{C}_t = \mathbf{Q}^T\mathbf{V}_{(t)}\mathbf{R}^{-1}$.

Matrix $\mathbf{C}_t$ is more useful in computing the Hessian. We can show that $g_{t,s}$ in equation (2.4) is identical to

$$g_{t,s} = -\sum_{i=1}^{P}\sum_{j=1}^{P} c_{t,ii}\ c_{s,jj} + \sum_{i=P+1}^{N}\sum_{j=1}^{P} c_{t,ij}\ c_{s,ij}.$$

That is, to obtain $g_{t,s}$, the elements of $\mathbf{C}_t$ below the Pth row are multiplied by the corresponding elements of $\mathbf{C}_s$ and the sum of the products is formed. From this, the sum of the products of the elements of $\mathbf{C}_t$ in the first P rows with the corresponding elements in $\mathbf{C}_s^T$ is subtracted. Hence the approximate Hessian can be computed by

$$\begin{aligned}&\frac{\partial^2|\mathbf{V}^T\mathbf{V}|}{\partial q_t \partial q_s} \\ &= 4\,|\mathbf{V}^T\mathbf{V}|\left(\sum_{i=1}^{P} c_{t,ii}\right)\left(\sum_{j=1}^{P} c_{s,jj}\right) \qquad (3.1) \\ &\quad + 2\,|\mathbf{V}^T\mathbf{V}|\left(-\sum_{i=1}^{P}\sum_{j=1}^{P} c_{t,ij}\,c_{s,ji} + \sum_{i=P+1}^{N}\sum_{j=1}^{P} c_{t,ij}\,c_{s,ij}\right)\end{aligned}$$

Equation (3.1) permits very efficient evaluation of the Hessian, because once the QR decomposition of $\mathbf{V}$ is done and the matrices $\mathbf{C}_t$, $t = 1, 2, \cdots, P$ are formed, it is only necessary to collect a few inner products. Although $\mathbf{Q}^T$ occurs as a factor in $\mathbf{C}_t$, the matrix $\mathbf{Q}$ is not explicitly formed; instead, a product such as $\mathbf{Q}^T\mathbf{V}_{(t)}$ is formed by applying Householder transformations to $\mathbf{V}_{(t)}$ (Dongarra et al., 1979).

4. A Concluding Remark

Applying the QR decomposition in nonlinear experimental design was core part of this thesis. At first, we derive the gradient and the Hessian of the determinant criterion, which is the D-optimal design criterion for the precise estimation of the model parameters. We showed how effienctly these quantities and the objective function can be computed using the QR decomposition of the derivative matrix $\mathbf{V}$.

For the future study, the first work should be the implementation of the computing algorithms developed in this thesis. Although a few experimental design software for the linear model are being used, none have been proposed for the nonlinear model. It may be possible to find the values of $\mathbf{x}_n$ which maximize $|\mathbf{V}^T\mathbf{V}|$ without using the derivatives. However, utilizing the gradient and Hessian would help to speed up the iteration procedure and to locate the exact optimal point. Whatever method we use, there is an inherent difficulty in implementing an optimization algorithm for experimental design, which is due to the interchangeability of the design points. For example, if (x_1^*, x_2^*) is an optimal point, then so is (x_2^*, x_1^*). To handle this problem, we may have to impose some constraints on the design space such as $a \le x_1 \le x_2 \le b$. Since constrained optimization is usually more difficult than unconstrained one, a way to avoid this kind of constraints would be desirable.

References

1. Atkinson, A. C. and V. V. Fedorov, *Optimal design: Experiments for discriminating between several models.*, Biometrika **62(2)** (1975), 289-304.

2. Bates, Douglas M. and Donald G. Watts, *Multiresponse estimation with special application to system of linear differential equation(with disscussion).*, Technometrics **27(4)** (1985), 329-360.
3. Box, G. E. P., and H. L. Lucas, *Design of experiments in nonlinear situations.*, Biometrika **46** (1959), 77-90.
4. Box, M. J., *An experimental design criterion for precise parameter estimation of a subset of the parameters in a nonlinear model.*, Biometrika **58** (1971), 149-153.
5. Cochran W. G., *Experiment for nonlinear functions.*, Journal of the American Statistical Assocation **68** (1973), 771-778.
6. Donggarra J. J., J. R. Bunch, C. B. Moler, and G. W. Stewart, *Linpack Uesr's Guide.*, SIAM, Philadelpia, 1979.
7. Draper, N. R. and W. G. Hunter, *The use of prior distribution in the design of experiments for parameter estimation in nonlinear situations*, Biometrika **54** (1967), 147-153.
8. Fedorov, V. V. (Translated by W. J.Studden and E. M.Klimko), *Theory of optimal experiments.*, Academic Press., New York, 1972.
9. Golub, G. H. and V. Pereyra, *The differentation of psedo-inverses and non-linear least squares problems whose variables separate.*, Journal of SIAM **10** (1973), 413-432.
10. Hill, W. J., W. G. Hunter, and D. W. Wichern, *A joint design criterion for the dual problem of model discrimination and parameter estimation.*, Technometrics **10** (1968), 145-160.
11. Searl, S. R., *Matrix algebra: Useful for statistics*, John Wiley and Sons, New York, 1982.

Department of Computer Science and Statistics,
Han-seo University,
Chungnam 356-820, Korea