Lecture Notes in Mathematics

Edited by A. Dold and B. Eckmann

1005

Numerical Methods

Proceedings of the International Workshop
Held at Caracas, June 14 – 18, 1982

Edited by V. Pereyra and A. Reinoza

Springer-Verlag
Berlin Heidelberg New York Tokyo 1983

CONTINUATION METHODS AND PARAMETRIZED NONLINEAR LEAST SQUARES: TECHNIQUES AND EXPERIMENTS

Ralph Baker Kearfott
Department of Mathematics and Statistics
University of Southwestern Louisiana
Lafayette, Louisiana 70504

1. Introduction

Suppose $H : R^{n+1} \to R^n$, and suppose the n by $n+1$ Jacobi matrix $J(H)(y)$ is nonsingular over the set $\mathcal{S} = \{y \mid H(y) = \theta_n\}$, so that $\mathcal{S}$ is equal to a disjoint union of arcs and circles.

Numerically tracing the arcs and circles in $\mathcal{S}$ has been the subject of much study (cf. e.g. [11] pp. 230-239, [1], and [5] for surveys). One might put applications into two broad categories: (1) models of systems in which the (n+1)-st parameter occurs naturally, and in which a so-parametrized family of solutions is desired, and (2) systems of equations in which the (n+1)-st parameter is introduced artificially, as an aid to solving a system $F(x) = \theta_n$, where $F : R^n \to R^n$. For example, $H(x,t)$, $x \in R^n$, $t \in R$, may be defined by:

$$H(x,t) = F(x) - (1-t)F(x^{(0)}), \tag{1.1}$$

for some $x^{(0)} \in R^n$; the problem of finding x^* such that $F(x^*) = \theta_n$ is then solved by following the path $H(x,t) = \theta_n$ starting at $(x^{(0)}, 0)$ and ending at $(x^*, 1)$. In theory, this technique is always applicable when $J(H)(y)$ is nonsingular over R^n (cf. [11], p. 231).

Recently, such homotopy, or continuation methods have been applied to nonlinear least squares problems (cf. e.g. [12]), where $F : R^n \to R^m$, $m \geq n$, and a least squares solution of $F(x) = \theta_m$ is desired. For example, the Newton homotopy (1.1) may be used, so that $H(x,t) : R^{n+1} \to R^m$, and a one-parameter family of problems is solved. Alternatively, the n by n Hessian matrix $\mathcal{H}(\phi)(x)$, where $\phi(x) = \frac{1}{2} F^T(x)F(x)$, may be numerically singular, so that it is appropriate to think of the least-squares solution to $F(x) = \theta_m$ as a curve in R^n. These cases correspond to the previously mentioned artificial homotopy and naturally occurring parametrized system, respectively.

The usual methods for solving non-parametrized nonlinear least squares problems include the Levenberg-Marquardt methods and implementations of quasi-Newton updates in conjunction with model trust regions or "doglegs" (cf. e.g. [3]). In these methods, the special structure of the nonlinear least squares problem is exploited.

In this paper, we describe some logical ways of incorporating nonlinear least squares structure in a general continuation method. In Section 2 we review our basic continuation method. In Section 3 we indicate how nonlinear least squares structure can be incorporated. In Section 4 we give some numerical results. We draw conclusions and indicate directions for further investigations in Section 5.

2. The Basic Method

The basic method is described in [9], [7], [8], and in previous works appearing as references therein, so we merely outline it here. Suppose $y(s) \subset R^{n+1}$ is a solution component, parametrized in terms of arclength s, of $H(y(s)) = \theta_n$, and suppose H is sufficiently smooth. Then:

$$J(H)(y(s))\dot{y}(s) = \theta_n, \quad \| \dot{y}(s) \|_2 = 1. \tag{2.1}$$

If, in addition, $y^{(0)} \in R^{n+1}$ is known with $H(y^{(0)}) = \theta_n$, then (2.1) and $y^{(0)}$ define an initial value problem which, in principle can be solved by any good O. D. E. software. We, however, prefer the predictor-corrector approach outlined here, since error control is localized, since it is easier to take advantage of structure in the problem, and since empirical results indicate it competes favorably in efficiency, in certain cases (cf. [7]).

Given $y^{(k)} \in R^{n+1}$ with $H(y^{(k)}) \doteq \theta_n$, we first find $b^{(k)}$ with $J(H)(b^{(k)}) \doteq \theta_n$, $\| b^{(k)} \|_2 = 1$, and we define:

$$z^{(k)} = y^{(k)} + \delta_k b^{(k)}, \tag{2.2}$$

for some predictor stepsize δ_k. We then correct $z^{(k)}$ in a hyperplane approximately perpendicular to $b^{(k)}$; that is, we take $y^{(k+1)}$ to be a solution to:

$$G(y^{(k+1)}) = \begin{bmatrix} H(y^{(k+1)}) \\ (b^{(k)})^T (y^{(k+1)} - z^{(k)}) \end{bmatrix} = \theta_{n+1}. \tag{2.3}$$

(See [7] for illustrations.)

It is sometimes advantageous not to simply compute the n by $n+1$ matrix $J(H)(y)$ at each point needed in (2.1), (2.2), and the iterative solution of (2.3), but to obtain approximations to it via, say, Broyden updates ([7]). In this context, $J(H)$ may be approximated reasonably well without substantially more evaluations of H by performing occasional special Powell correction steps (ibid.). For dense $J(H)$, the updates may be computed in $O(n^2)$ operations on a factored form of $J(H)$, and can be applied appropriately to sparse or banded systems.

3. The Nonlinear Least Squares Structure

As explained above, the (n+1)-st parameter in parametrized nonlinear least squares problems may occur naturally, artificially, or implicitly in problems with a singular Hessian matrix. For simplicity, let us assume here that $H : R^{n+1} \to R^m$, and that it is required, for fixed $t \in R$, to find a least squares solution over $x \in R^n$ of $H(x,t) = \theta_m$. (The case where $\mathcal{H}(\phi)$ is numerically singular, i.e., where the parametrization is implicit, will be treated elsewhere.) Then $\min_x \phi(x,t)$ occur at x for which:

$$J_x^T(H)(x,t)H(x,t) = \theta_n, \tag{3.1}$$

where $J_x(H)$ is the m by n Jacobi matrix of H with respect to the first n parameters x. Define:

$$\widetilde{H}(x,t) = J_x^T(H)(x,t)H(x,t) : R^{n+1} \to R^n. \tag{3.2}$$

The n by $n+1$ Jacobi matrix of $\widetilde{H}$ is then:

$$J(\widetilde{H})(x,t) = J_x^T(H)(x,t)J(H)(x,t) + S(x,t), \tag{3.3}$$

where $J(H)(x,t)$ is the full m by $n+1$ Jacobi matrix of H and where $S(x,t)$ is a matrix whose entries depend on the second-order partial derivatives of the components of H.

In our procedure, we will compute $J(H)$, and hence $J_x(H)$, "precisely" with finite differences, and we will approximate $S(x,t)$ with quasi-Newton updates. Specifically, suppose $y \in R^{n+1}$ and $y^+ = y + s$. (The vector s may represent a predictor step $\delta_k b_k$ as in (2.2), a corrector step in the iterative solution of (2.3), or a special Powell correction.) Suppose S (as in (3.3)) is the old approximate second-order part of the matrix $J(H)$, and denote the new approximate second-order part of $J(H)$ by S^+. Here, we will compute S^+ by performing a Broyden update:

$$S^+ = S + vs^T / \|s\|_2^2, \tag{3.4}$$

where

$$v = [J_x^T(H)(y^+) - J_x^T(H)(y)]H(y^+) \tag{3.5}$$

(cf. formula (3.2) in [4]).

The parametrized nonlinear least squares method will consist of the continuation method applied to $\tilde{H}$, with $J(\tilde{H})$ approximated as indicated.

Updates other than (3.4) which take better advantage of the natural scaling inherent in the first n parameters have been suggested ([2]). They will be reported on elsewhere.

4. Software Tools and Numerical Experiments

The numerical experiments appearing here were run using a preliminary version of a modular software package for solving general classes of general continuation and nonlinear least squares problems. In this package, the basic predictor-corrector method is defined in a controlling module, while all matrix operations and various other application-specific tasks are performed in external modules the names of which are passed to the controlling module as arguments. These tasks include adjustment of predictor stepsize, choice of predictor direction, detection of non-convergence of corrector iteration, computation of corrector iteration, computation of corrector iterates, detection of bifurcation points, and determination of directions and stepsizes for following arcs away from bifurcation points. Thus, the package provides a controllable environment for research into and comparison of various matrix updates, corrector iteration processes, adaptive stepsize control, bifurcation techniques, etc. Additionally, the package will be able to efficiently handle sparse or specially structured problems.

In our first experiments, $H : R^{n+1} \to R^n$ was defined artificially via (1.1). Two test problems have been published in this context in [12]; they are defined as follows, where $F(x) = (f_1(x), f_2(x), \ldots, f_m(x))^T$:

$$\begin{aligned} &m = 24, \quad n = 4, \quad u_k = .1(k - 1 \\ &\tilde{x} = (60.137,\ 1.371,\ 3.112,\ 1.761)^T \\ &\tilde{f}_k(x) = x_1 x_2^{u_k} \sin(x_3 u_k + x_4) \\ &f_k(x) = \tilde{f}_k(x) - \tilde{f}_k(\tilde{x}), \end{aligned} \tag{4.1}$$

and

$$m = 16, \quad n = 5, \quad u_k = .1(k - 1)$$

$$\tilde{x} = (53.81,\ 1.27,\ 3.012,\ 2.13,\ .507)$$

(4.2) $$\tilde{f}_k(x) = x_1 x_2^{u_k}[\tanh(x_3 u_k) + \sin(x_4 u_k)]\cos(u_k e^{x_5})$$

$$f_k(x) = \tilde{f}_k(x) - \tilde{f}_k(\tilde{x}).$$

In [12] four starting vectors $x^{(0)}$ were tried for (4.1) and $\sin x^{(0)}$ were tried for (4.2). The method employed there consisted of choosing, a priori, equally spaced t: $t_0 = 0$, $t_1 = 1/p, \ldots, t_{p-1} = (p-1)/p$, $t_p = 1$ ($p = 10$, $p = 20$, or $p = 40$). Then, (1.1) was solved as a non-parametrized problem in R^n for successive t_j, using the solution at the t_{j-1} level as the starting vector for the t_j level. (This corresponds to the "elevator" predictor with constant stepsize (cf. [14], pp. 310-311). The Newton-Raphson method (with explicitly computed Hessian matrices) and the Gauss-Newton method were used for corrector iteration; in both cases, steplengths were damped and line searches were employed.

Here we compare the methods in [12] to the continuation method outlined in Sections 2 and 3, using roughly the same tolerances as in [12]. Differences include parametrization with respect to arclength, adaptive stepsize control, and use of quasi-Newton updates. Additionally, we employed no line searches or corrector step bounds, since we felt such might be unnecessary in the presence of careful stepsize control. We did not do special Powell correction steps, since preliminary experiments indicated these did not appreciably affect the algorithmic behavior.

In [12] numbers of evaluations of ϕ, numbers of Jacobi matrix evaluations, and numbers of Hessian matrix evaluations are given, while numbers of evaluations of H were counted in our algorithm. We assume an evaluation of ϕ is equivalent to an evaluation of H, and we compare our evaluations of H to the "equivalent evaluations" given in [12].

The results for (4.1) and (4.2) appear in Table 4.1 and Table 4.2, respectively. In each, the first column gives the starting vector $x^{(0)}$, the second (NEQV) gives the number of equivalent evaluations from [12], and the third column (NEV) gives our number of evaluations of H.

Table 4.1 does not include results for our method for the second starting point. This is because the arc diverges to ∞, (See Figure 4.1 where a graph of x_2 versus t appears for this starting point.) We assume success was reported in [12] because the corrector iteration scheme was a descent method, and the corrector iterates jumped to another arc somewhere between $t = .35$

and $t = .4$.

In [12] a relative tolerance of 10^{-2} was used for determining convergence of corrector iteration at intermediate points, while a relative tolerance of 10^{-6} was used at the $t = 1$ hyperplane. We used tolerances of 10^{-4} and 10^{-6}, respectively, since these stricter tolerances actually led to less function evaluations. These are not strictly comparable, however, since we checked both the domain and range, and we used more sophisticated scaling.

In [12] more than one try, with different parameters, was reported for each entry in Table 4.1 and Table 4.2. We reported only those tries most favorable to the method in [12], though there was a wide variation.

5. Conclusions

We have presented a continuation method for nonlinear least squares using quasi-Newton updates, careful adaptive stepsize control, and parametrization with respect to arc length; we did not use model trust regions, line searches, or damping in corrector iterations. Such a method can be viewed as an alternate procedure for globalizing the algorithm; we feel the stepsize control and efficiency vis-a-vis alternate methods should be further investigated.

The numerical comparisons were somewhat disappointing with regard to total number of function evaluations. However, we feel the arcs were reliably followed, and indications are that improvements in stepsize control will make a large difference. Furthermore, the problems in [12] were zero-residual problems at the $t = 1$ hyperplane, and the most efficient method reported in [12] was the Gauss-Newton method. The update technique described in Section 3 of this paper is specifically for problems with non-zero residuals, and should perform relatively better on such problems. In [12] difficulties were mentioned for problem (4.2), and a special parameter q, altering (1.1), was introduced. We encountered no such difficulties. We were also able to determine more of the structure of the arcs.

We have also done preliminary tests using the homotopy (1.1) and the 18 nonlinear least squares problems and starting points described in [10]. In several instances the arcs defined by (1.1) did not reach the $t = 1$ hyperplane, indicating that (1.1) cannot always be used naively (see also [13]). In other instances the approximate Hessian matrices became singular to working precision, indicating possible bifurcation points, or at least a need to use a more sophisticated corrector iteration. In other cases, the predictor stepsize went to zero, indicating a need for better stepsize control. Over half of the 54 cased tried, however, were successful with a reasonably small number of function evaluations.

Our experience with the general continuation method and with continuation methods for nonlinear least squares leads us to believe the nonlinear least square problems are inherently more difficult. Better use of the natural scaling in the first n variables may lead to updates better able to handle such ill-conditioned problems ([2]).

The method holds promise for parametrized nonlinear least squares problems where the parameter occurs naturally. Results concerning this will be reported elsewhere.

References

1. Allgower, E., and Georg, K., "Simplicial and continuation methods for approximating fixed points and solutions to systems of equations," SIAM Review 22 no. 1, pp. 28-85, 1980.

2. Dennis, J. E., "An update for parametrized nonlinear least squares," private communication, 1982.

3. Dennis, J. E., and Schnabel, R. B., Numerical Methods for Unconstrained Optimization and Nonlinear Least Squares, to appear, 1982.

4. Dennis, J. E., Gay, D. M., and Welsch, R. E., "An adaptive nonlinear least-squares algorithm," Technical Summary Report no. 2010, Mathematics Research Center, University of Wisconsin, 1979.

5. Garcia, C. B., and Zangwill, W. I., Pathways to Solutions, Fixed Points, and Equilibria, Prentice-Hall, 1981.

6. Georg, K., "On tracing an implicitly defined curve by quasi-Newton steps and calculating bifurcation by local perturbations," SIAM Journal on Scientific and Statistical Computing 2 no. 1, pp. 35-50, 1981.

7. Kearfott, R. B., "A derivative-free arc continuation method and a bifurcation technique," Springer Lecture Notes no. 878, 1981.

8. Kearfott, R. B., "Some general bifurcation techniques," to appear in the SIAM Journal on Scientific and Statistical Computing.

9. Li, T. Y., and Yorke, J. A., "A simple reliable numerical algorithm for following homotopy paths," in Analysis and Computation of Fixed Points, ed. S.M. Robinson, Academic Press, 1979.

10. More', J. J., Garbow, B. S., and Hillstrom, K. E., "Testing unconstrained optimization software," ACM Transactions on Mathematical Software 7 no. 1, pp. 17-41, 1981.

11. Ortega, J. M., and Rheinboldt, W. C., Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, 1970.

12. de Villiers, N., and Glasser, D., "A continuation method for nonlinear regression," SIAM Journal on Numerical Analysis 18 no. 6, pp. 1139-1154, 1981.

13. Watson, L. T., "Engineering applications of the Chow-Yorke Algorithm," Applied Mathematics and Computing, 1981.

$x^{(0)}$	NEQV	NEV
(1.0, 8.0, 4.0, 4.412)	672	3324
(1.0, 8.0, 8.0, 1.0)	758	----
(1.0, 8.0, 1.0, 4.412)	263	1464
(1.0, 8.0, 4.0, 1.0)	583	1980

Table 4.1: Results for function (4.1).

$x^{(0)}$	NEQV	NEV
(45.0, 2.0, 2.5, 1.5, .9)	197	1008
(41.0, .8, 1.4, 1.8, 1.0)	307	1463
(45.0, 2.0, 2.1, 2.0, 0.9)	197	952
(45.0, 2.5, 1.7, 1.0, 1.0)	405	2051
(35.0, 2.5, 1.7, 1.0, 1.0)	381	1764
(42.0, 0.8, 1.8, 3.15, 1.0)	977	5054

Table 4.2: Results for function (4.2).