RANK-DEFICIENT AND ILL-CONDITIONED NONLINEAR LEAST SQUARES PROBLEMS *

C. T. KELLEY†, I. C. F. IPSEN†, AND S. R. POPE‡

Abstract. In this paper we review classical results by Ben-Israel, Boggs, and Tanabe on algorithms for rank-deficient nonlinear least squares problems and compare them to our recent perturbation theory results for such problems. Using a simple example, we show how sufficiently large perturbations can lead to poor convergence or even incorrect results, and how the application of subset selection can solve some of these problems.

Key words. Nonlinear least squares, Rank-Deficient Jacobian, Perturbation Theory

AMS subject classifications. 15A18, 65F20, 65F22, 65F15, 65F35, 65H10, 65K10, 65L09

1. Introduction. Rank-deficient nonlinear least squares problems arise when nonlinear models have more free parameters than the model can resolve or when the parameters are dependent in ways that are difficult for the modeler to detect. Such problems occur, for example, in the life sciences [6, 10, 16, 17], which provided the motivation for our work in [11].

If the Jacobian is Lipschitz continuous and has constant rank, then classical results [1, 2, 13, 18, 19] say that the Moore-Penrose inverse [3] of the Jacobian can be used in a Gauss-Newton or Levenberg-Marquardt iteration to converge to the manifold of solutions, at least in exact arithmetic. The perturbation theory in our recent paper [11] shows that this algorithm also works well if the rank of the Jacobian can be clearly identified, *i. e.* there is a sufficiently large gap in the computed singular values between the smallest singular value of interest and the rest, and the nonlinear residual and Jacobian are computed sufficiently accurately.

If the errors in the computed residual and Jacobian are too large for the classical results to hold, then the algorithm can fail to converge [17] or, as we will see in § 3, produce incorrect results. One remedy is to use subset selection [7–9] to identify a useful subset of the parameters, set the remainder to nominal values, and solve a smaller problem for the subset. One would hope that the smaller problem has a full-rank Jacobian and is well-conditioned.

In § 2 we discuss the nonlinear least squares problem of interest in this paper and our assumptions on the rank-deficiency. We then state one version of the classical results and our recent perturbation result. The perturbation analysis identifies conditions in which the results can be poor in practice, and we discuss those possibilities. Finally in § 3 we present a variation of the study from [11].

2. Theory. We consider a class of rank-deficient nonlinear least squares problems of the form

$$\min \|R(p)\|^2/2 \tag{2.1}$$

where the norm is the Euclidean norm, all minimizations are over $\mathbb{R}^N$, $R : \mathbb{R}^N \to \mathbb{R}^M$, $M > N$, and the rank of the Jacobian R' is $k < N$. We assume that the rank of $R'(p)$ is independent of p.

2.1. Results in Exact Arithmetic. The classical Gauss-Newton [5, 13] method for nonlinear least squares problems updates a current approximation to the optimal point p_c by adding the Gauss-Newton step s which solves the linearized problem

$$\min \|R'(p_c)s + R(p_c)\|^2, \tag{2.2}$$

and then updates p_c to a new iteration

$$p_+ = p_c + s.$$

*This work has been partially supported by National Science Foundation Grant DMS-0707220 and Army Research Office Grants W911NF-06-1-0412, and W911NF-07-1-0112.

†North Carolina State University, Department of Mathematics, Box 8205, Raleigh, NC 27695-8205, USA (ipsen@ncsu.edu,Tim_Kelley@ncsu.edu)

‡SAS Institute, 100 SAS Campus Drive, Cary, NC 27513, USA (scott.pope@gmail.com)

In the full rank $k = N$ case, we obtain the standard Gauss-Newton formula

$$s = -(R'(p_c)^T R'(p_c))^{-1} R'(p_c)^T R(p_c).$$

If $k < N$, then (2.2) does not have a unique solution. The natural thing to do is take the minimum norm solution of (2.2)

$$s = -R'(p_c)^\dagger R(p_c),$$

where $R'(p_c)^\dagger$ is the Moore-Penrose [8] inverse of R'.

The Gauss-Newton model Hessian will be singular if R' is rank-deficient and this will create problems in convergence if the initial iterate is far from the solution. A remedy for this is to add a multiple of the identity to obtain the Levenberg-Marquardt [5, 13–15] iteration

$$s = (\nu_c I + R'(p_c)^T R'(p_c))^{-1} R'(p_c)^T R(p_c).$$

Here ν_c is the Levenberg-Marquardt parameter at the current point. The Levenberg-Marquardt parameter ν may be held constant or varied as the iteration progresses.

The classical results describe the convergence of the iteration for certain rank-deficient problems. We will examine a subset of that class in this work, and present our assumptions before stating a version of the classical results.

In [11] we assumed that R could be factored as the composition of two Lipschitz continuously differentiable functions $\tilde{R}$ and B. We formalize this as

ASSUMPTION 2.1. *Assume that*

$$R(p) = \tilde{R}(B(p)), \tag{2.3}$$

where $\tilde{R}' : \mathbb{R}^k \to \mathbb{R}^M$ has full column rank k and $B' : \mathbb{R}^N \to \mathbb{R}^k$ has full row rank k, and the smallest singular value of $\tilde{R}$ and smallest non-zero singular value of B are bounded away from zero.

The idea is that the nonlinear model will receive N parameters, but the internal dependencies in the parameters or lack of sensitivity lead to only k fittable parameters. We express this for theoretical purposes by using the reduction function B. The reader should keep in mind that the solvers only see R and that the factors $\tilde{R}$ and B are only used in the analysis. Assumption 2.1 implies that the set of local minimizers of $\|\tilde{R}\|^2$ consists of isolated points. Moreover, if b^* is a local minimizer of $\tilde{R}$ and $\|\tilde{R}(b^*)$ is sufficiently small, then the Gauss-Newton iteration will be locally convergent.

For convenience we set

$$f(p) = \|R(p)\|/2.$$

We will assume for convenience that b^* is the unique minimizer of $\|\tilde{R}\|$ and let

$$\mathcal{Z} = \{p \,|\, f(p) = f^*\} = \{p \,|\, B(p) = b^*\}. \tag{2.4}$$

Following [13] we define

$$d(p) = \min_{z \in \mathcal{Z}} \|z - p\|.$$

We seek algorithms that drive $d(p)$ to zero.

Our regularity assumptions on $\tilde{R}$ and B imply that R' has rank $k < N$ everywhere, and hence the smallest singular value of R' is bounded away from zero on any compact set.

As an example we state the result from [1] for the Gauss-Newton iteration for the special case where Assumption 2.1 holds. Similar results for damped the Levenberg-Marquardt iteration (with fixed Levenberg-Marquardt parameter) are reported in [2, 18, 19].

THEOREM 2.1. *Let Assumption 2.1 hold and assume that b^* is the unique minimizer of $\tilde{R}$. If $\|\tilde{R}(b^*)\|$ is sufficiently small and p_0 is sufficiently near $\mathcal{Z}$ then the regularized Gauss-Newton iteration*

$$p_{n+1} = p_n - R'(p_n)^\dagger R(p_n) \tag{2.5}$$

converges to $p^* \in \mathcal{Z}$. Moreover

$$d(p_{n+1}) = O(d(p_n)^2 + \|\tilde{R}(b^*)\|d(p_n)). \tag{2.6}$$

In [11] we extended the classical results to include a standard trust-region approach to managing the Levenberg-Marquardt parameter [4, 5, 13]. In this approach we take a trial Levenberg-Marquardt step from p_c to obtain

$$s_t = -(\nu_c I + R'(p_c)^T R'(p_c))^\dagger R'(p-c)^T R(p_c)$$

Then we compute the actual reduction in the objective function

$$ared = f(p_c) - f(p_c + s_t),$$

and compare it to the reduction predicted by the quadratic model

$$pred = -\nabla f(p_c)^T s_t/2 = -(R'(p_c)^T R(p_c))^T s_t/2.$$

The algorithm which decides to accept or reject the step and to decrease or increase ν_c is standard. One begins with several parameters:

$$0 < \omega_{down} < 1 < \omega_{up}, \quad \nu_0 \geq 0, \qquad \text{and} \qquad 0 \leq \mu_0 < \mu_{low} \leq \mu_{high} < 1.$$

Typical values, which we use in the computations we report in this section are

$$\mu_0 = 10^{-4}, \quad \mu_{low} = 1/4, \quad \mu_{high} = 3/4, \quad \omega_{down} = 1/2, \quad \text{and} \quad \omega_{up} = 2. \tag{2.7}$$

We express the outcome interms of the current values of the point and the Levenberg parameter p_c and ν_c and, if the step is successful the new point and new Levenberg parameter p_+ and ν_+. The decision process is:

- If $ared/pred < \mu_0$ then set $\nu_c = \max(\omega_{up}\nu_c, \nu_0)$, keep the current point, and recompute the trial point with the new value of ν_c.
- If $\mu_0 \leq ared/pred < \mu_{low}$, then accept the step and set $p_+ = p_t$ and $\nu_+ = \max(\omega_{up}\nu_c, \nu_0)$.
- If $\mu_{low} \leq ared/pred$, then accept the step and set $p_+ = p_t$ and $\nu_+ = \nu_c$.
  If $\mu_{high} < ared/pred$, then set $\nu_+ = \omega_{down}\nu_c$.
  If $\nu_c < \nu_0$ then set $\nu_+ = 0$.

We will refer to the Levenberg-Marquardt iteration with the algorithm outlined above for managing ν as the **standard Levenberg-Marquardt** algorithm.

In the full rank case [4, 5] one can show that if f has bounded level sets, then

$$\liminf \|\nabla f(p_n)\| = 0.$$

In particular, $\limsup \nu_n < \infty$. In the rank-deficient case, it is less clear that the Levenberg-Marquardt parameter remains bounded. In [11] we proved a local convergence result that includes boundedness of ν.

THEOREM 2.2. *Let Assumption 2.1 hold and assume that b^* is the unique minimizer $\tilde{R}$. The if $\|\tilde{R}(b^*)\|$ is sufficiently small and p_0 is sufficiently near $\mathcal{Z}$ then $\limsup \nu_n < \infty$ and the standard Levenberg-Marquardt Gauss-Newton iteration converges to $p^* \in \mathcal{Z}$.*

Theorem 2.2 does not have a convergence estimate like (2.6) because we were unable to show in that paper that $\nu_n \to 0$.

2.2. Perturbation Theory and the Truncated SVD Method. If there are errors in the evaluation of R, then the computed R may have N non-zero singular values, even though R itself has only k. If one has knowledge of the correct value of k or there is a clear gap in the the singular values of the computed R, one can simply replace the $N - k$ smallest singular values by zero. In [11] we considered errors in R' and not R. The reason for this is that errors in R can be considered to be errors in the data. However, errors in the data can create failure in the iteration, as we will see in § 3.2.

We assume that we have an error E in the computation of R' and let

$$J = R' + E \tag{2.8}$$

be the computed Jacobian. Suppose we know that R' has exactly k non-zero singular values. We will then take the SVD of J

$$J = U\Sigma V^T$$

where

$$\Sigma = diag(\sigma_1, \ldots, \sigma_k, \sigma_{k+1}, \ldots \sigma_N).$$

We let

$$\tilde{\Sigma} = diag(\sigma_1, \ldots, \sigma_k, 0, \ldots 0),$$

and define

$$\tilde{J} = U\tilde{\Sigma}V^T. \tag{2.9}$$

Our **truncated SVD Levenberg-Marquardt** iteration is the same as the standard iteration except that we replace R' by $\tilde{J}$ throughout. The perturbation result compares the truncated SVD Levenberg-Marquardt step for the computed R'

$$\tilde{s} = -(\nu + \tilde{J}^T\tilde{J})^\dagger \tilde{J}^T R(p) \tag{2.10}$$

with the standard Levenberg-Marquardt step for an exact computation of R'

$$s_t = -(\nu + R'(p)^T R'(p))^\dagger R'(p)^T R(p).$$

The bound on $\|\tilde{s} - s_t\|$ is technical and depends on bounds for the singular values. Assumption 2.1 implies that the k non-zero singular values of R' are bounded from below by $\bar{\sigma}_k > 0$ and from above by $\bar{\sigma}_1$. For $\nu \geq 0$ define

$$\eta(\nu) = \max_{\bar{\sigma}_k \leq \sigma \leq \bar{\sigma}_1} \frac{\sigma}{\nu + \sigma^2}. \tag{2.11}$$

THEOREM 2.3. *Let Assumption 2.1 hold and assume that b^* is the unique minimizer of $\tilde{R}$. Let $k < N$ and let J, E and $\tilde{J}$ be given by* (2.8) *and* (2.9)*. If*

$$\theta\|E\|_F < 1, \text{ where } \theta = \frac{2}{\bar{\sigma}_k - 2\|E\|} < 1,$$

then

$$\|s_t - \tilde{s}\| \leq \left[2\eta(\nu)\|\tilde{s}\| + \left(\frac{1}{\nu + \bar{\sigma}_k^2} + 2\eta(\nu)\theta\right)\|\tilde{R}\|\right]\|E\|_F + \omega\|E\|_F^2, \tag{2.12}$$

where

$$\omega = \frac{\|\tilde{s}\|}{\nu + \bar{\sigma}_k^2} + 2\eta(\nu)\theta^2\|\tilde{R}\|.$$

One consequence of the theorem is that if p is sufficiently near $\mathcal{Z}$, $\|E\|$ and $\|\tilde{R}(b^*)\|$ are sufficiently small, then $\tilde{s}$ is a good approximation to s_t, and one should expect the numerical observations to conform to the theory. The condition

$$\theta\|E\| = \frac{2\|E\|_F}{\bar{\sigma}_k - 2\|E\|} < 1 \tag{2.13}$$

requires that E be small enough so that there is a clear gap between the smallest nonzero singular value of R' and the $N-k$ smaller singular values of J. The presence of $\|E\|_F^2$ on the right of (2.12) limits the accuracy of the approximation to the range $\|E\|_F = O(\|\tilde{s}\|)$, which is a common limit on the resolution of the solution of a nonlinear equation [12] if the error in the Jacobian is no larger than the square root of the error in the evaluation of the residual (which for us would be $\|\tilde{R}(b^*)\|$). So, while $\|E\|_F = O(\|\tilde{s}\|)$, we obtain

$$\|s_t - \tilde{s}\| \leq O(\|\tilde{s}\|^2 + \|\tilde{s}\|\|R\|), \tag{2.14}$$

which is the same order as the error term in (2.6) if p is sufficiently near $\mathcal{Z}$.

On the other had, as we will see in the examples, the truncated SVD Levenberg-Marquardt method is very sensitive to large $\|\tilde{R}(b^*)\|$.

2.3. Subset Selection. Subset selection chooses k linearly independent columns from the Jacobian. In the context of this paper, this amounts to identifying the "best" k of the N original set of design parameters, setting the remaing $N-k$ to nominal values, and then solving for the smaller design vector. The result of this is that the nonlinear least squares problem we actually solve has a Jacobian of full rank (k).

Subset selection on the matrix R' produces a permutation matrix Π, which brings the k columns J_1 to the front, i.e. $R'\Pi = \begin{pmatrix} J_1 & J_2 \end{pmatrix}$. We use the strong rank revealing QR (SRRQR) algorithm by Gu and Eisenstat [9, Algorithm 4] in the exact version (with parameter $f = 1$). This SRRQR algorithm assures, among other things, for the singular values $\sigma_i(J_1)$ of J_1 that

$$\frac{\sigma_i}{\sqrt{1+k(N-k)}} \leq \sigma_i(J_1) \leq \sigma_i, \qquad 1 \leq i \leq k. \tag{2.15}$$

The reader should keep in mind that there need not be a unique way to select the columns. For the example in § 3 subset selection could select columns $(1,2,4)$ or $(1,3,4)$.

We assume that $\|E\|$ is small enough so that $J = R' + E$ has rank at least k and that subset selection picks the same set of k columns (as we must if we are to compare results directly). Hence

$$J\Pi = \begin{pmatrix} \tilde{J}_1 & \tilde{J}_2 \end{pmatrix},$$

where $\tilde{J}_1$ has k columns.

In either case, the step is now in $\mathbb{R}^k$, and we can express it by

$$s_t = (\nu I + J_1 J_1^T)^{-1} J_1^T R, \tag{2.16}$$

and for the perturbed case

$$\hat{s} = (\nu I + \tilde{J}_1 \tilde{J}_1^T)^{-1} \tilde{J}_1^T R. \tag{2.17}$$

The perturbation theory is now more favorable.

THEOREM 2.4. *Let Assumption 2.1 hold and assume that b^* is the unique minimizer of $\tilde{R}$. Let $k < N$ and let J, E and $\tilde{J}$ be given by* (2.8) *and* (2.9)*. Let subset selection be applied to R and J in such a way that the same k columns are selected and* (2.15) *holds. Let s_t and $\hat{s}$ be given by* (2.16) *and* (2.17)*. Then*

$$\|\tilde{s} - \hat{s}\| \leq \left[\tilde{\eta}(\nu)\|\tilde{s}\| + \frac{\|R\|}{\nu + \bar{\sigma}_k^2/(1+k(N-k))} \right] \|E\|,$$

where

$$\tilde{\eta}(\nu) = \max_{\bar{\sigma}_k/\sqrt{1+k(N-k)} \leq \sigma \leq \bar{\sigma}_1} \frac{\sigma}{\nu + \sigma^2}.$$

3. Example. As an example we consider a parameter identification problem for a driven harmonic oscillator. The equation is the the second order initial value problem

$$my'' + cy' + k_0 y = A\sin(\omega t), \qquad y(0) = y_0, y'(0) = y_0'. \tag{3.1}$$

Here m is the mass, c is the damping coefficient, and k_0 is the spring constant. We integrate the initial value problem from $t = 0$ to $t = 10$. The data are samples of the exact solution at the 100 equally spaced points $t = j/10$, $1 \le j \le 100$.

We seek to identify c, k_0, and a small perturbation of the mass. Think of a unit mass with an insect sitting on it; our task is to weigh that insect. In particular, we set $m = 1+10^{-3}\delta_m$ and seek to find δ_m. As it stands the nonlinear least squares problem would not be rank deficient. We will add the effects of modeling error by assuming that the modeler has replaced the damping constant c by a sum $c_1 + c_1$. Hence our model equation is

$$(1 + 10^{-3}\delta_m)y'' + (c_1 + c_2)y' + k_0 y = A\sin(\omega t),\ y(0) = y_0, y'(0) = y_0'. \tag{3.2}$$

The parameter vector is $p = (\delta_m, c_1, c_2, k_0)^T \in \mathbb{R}^4$. So in this example, $N = 4$, $M = 100$, and $k = 3$. If we let $y(t : p)$ be the solution of (3.2) with parameter vector p and let p^* be the "exact data", then the residual $R \in \mathbb{R}^{100}$ has components

$$R_i(p) = y(t_i : p) - y(t_i : p^*).$$

We will compute the columns of R' by solving the sensitivity equations. The ijth entry of R' is

$$\partial y(t_i : p)/\partial p_j$$

and we can compute $\partial y/\partial p_j$ by differentiating (3.2) with respect to p_j. For example, if we differentiate (3.2) with respect to $p_2 = c_1$ we see that $w_2 = \partial y/\partial c_1$ is the solution of the linear initial value problem

$$(1 + 10^{-3}\delta_m)w'' + (c_1 + c_2)w' + k_0 w = -y',\ y(0) = y'(0) = 0,$$

which is also the equation satisfied by $w_3 = \partial y/\partial c_2$.

In the computations for this section we solved a coupled system for the solution y and the four sensitivities. We converted the five second-order equations into a first-order system of ten equations, and solved that with the MATLAB `ode15s` integrator. In all our experiments we set

$$p^* = (1.23, 1, 0, 1)^T$$

and varied the tolerances for the integration and noise in the data to illustrate the theory from § 2.

We assign equal relative and absolute error tolerances to `ode15s.m`.

$$atol = rtol = \tau_{ivp}. \tag{3.3}$$

The tolerance for the initial value problem determines both the termination criterion for the nonlinear least squares solver and, most of the time, the number of columns requested from subset selection. We terminate the least squares iteration when the Levenberg-Marquardt iterates $\{p_n\}_{n\ge 0}$ satisfy

$$\|\nabla f(p_n)\| = \|J(p_n)^T R(p_n)\| \le 10\tau_{ivp} \qquad \text{or} \qquad |f(p_n) - f(p_{n-1})| < 100\tau_{ivp}^2. \tag{3.4}$$

Here J is the computed approximation of R'.

In each of the subsections that follow we report the results of parameter fits both with the truncated SVD approach where we discard the smallest singular value and a subset selection approach where we determine that there are only three non-zero singular values at the beginning of the optimization, identify three parameters, and fit only those, setting the other to zero.

In all cases $p_0 = \begin{pmatrix} 0 & 1 & 1 & .3 \end{pmatrix}^T$.

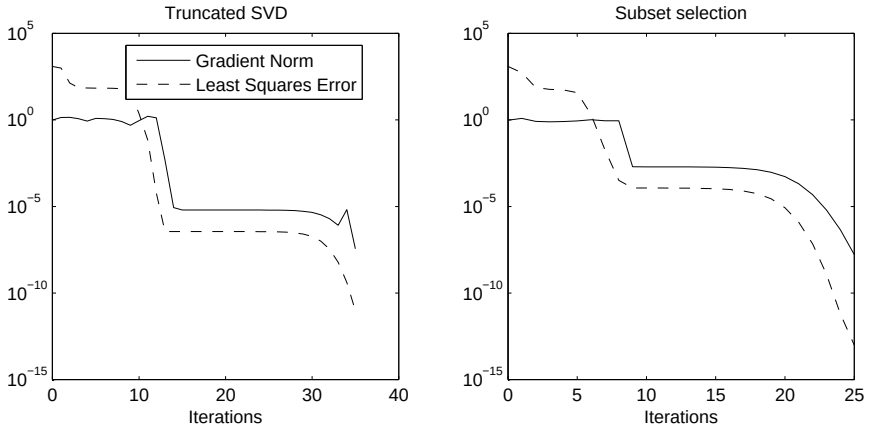
3.1. High-Accuracy Integration and Data. Here we show that the performance of the iteration and the quality of the results are good if (1) the residual is small and (2) the Jacobian is evaluated accurately. In this case we set

$$\tau_{ivp} = 10^{-8}$$

and use the values of the analytic solution for the data.

For this example both methods converge to the correct solution. The truncated SVD iteration converged to $(1.22, .5, .5, 1.0)^T$, which is on the manifold. The version where subset selection eliminated c_1 and converged to $(1.23, 1, 1)^T$ is a slightly better result. Figure 3.1 shows that both iterations converge rapidly, with subset selection about 20% faster.

FIG. 3.1. *Forced Oscillator: High Accuracy*

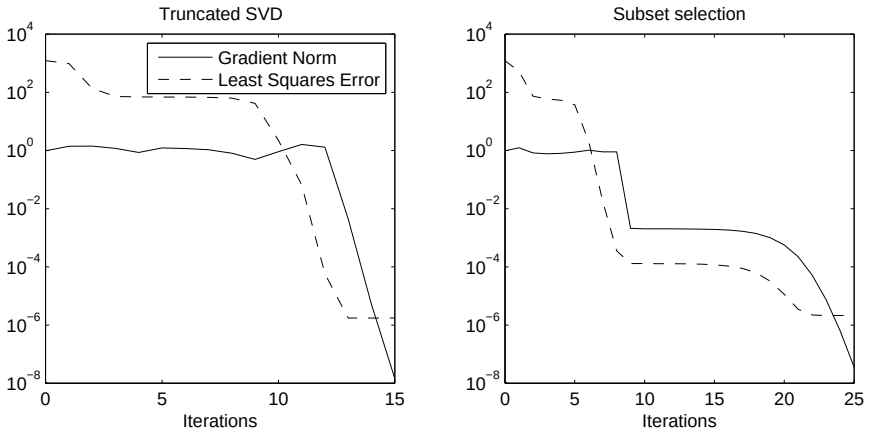


3.2. High-Accuracy Integration and Perturbed Data. In this example, taken from [11], we continue to use tight tolerances, but perturb the data by $1 + 10^{-4}r$, where r is a uniformly distributed random vector obtained with the MATLAB `rand` command.

This is a small residual, but not a zero-residual problem, which causes problems for the truncated SVD method. The truncated SVD method completely misses δ_m, obtaining a final result of $(.091, .5, .5, .998)^T$. Subset selection, on the other hand, obtained a reasonable fit of $(1.28, 1, 1)^T$.

The iteration statistics in Figure 3.2 show that subset selection seems to be converging more rapidly in the terminal phase, as one would expect from a full-rank problem.

FIG. 3.2. *Forced Oscillator: Noisy Data*



4. Conclusions. There is a classical local convergence theory for rank-deficient nonlinear least squares problems. We show that this theory predicts the outcome of computations well if the errors in the residual and Jacobian are small relative to the smallest non-zero singular value of the Jacobian. We show how subset selection can make the computations more robust if the errors are large and present computational evidence.

5. Acknowledgments. One of the authors (CTK) is very grateful to the organizers of the Sixth East Asia SIAM Conference for their invitation to speak at the conference and to submit this paper to the conference proceedings.

REFERENCES

[1] A. Ben-Israel, *A Newton-Raphson method for the solution of systems of equations*, J. Math. Anal. Appl., 15 (1966), pp. 243–252.
[2] P. T. Boggs, *The convergence of the Ben-Israel iteration for nonlinear least squares problems*, Math. Comp., 30 (1976), pp. 512–522.
[3] S. L. Campbell and C. D. Meyer, *Generalized Inverses of Linear Transformations*, Dover Press, New York, 1991.
[4] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust Region Methods*, no. 1 in MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2000.
[5] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, no. 16 in Classics in Applied Mathematics, SIAM, Philadelphia, 1996.
[6] L. M. Ellwein, S. R. Pope, A. Xie, J. Batzel, C. T. Kelley, and M. S. Olufsen, *Modeling cardiovascular and respiratory dynamics in congestive heart failure*. preprint, 2009.
[7] G. Golub, V. Klema, and G. Stewart, *Rank degeneracy and least squares problems*, Tech. Rep. STAN-CS-76-559, Computer Science Department, Stanford University, 1976.
[8] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, third ed., 1996.
[9] M. Gu and S. C. Eisenstat, *Efficient algorithms for computing a strong rank-revealing QR factorization*, SIAM J. Sci. Comput., 17 (1996), pp. 848–869.
[10] T. Heldt, *Computational Models of Cardiovascular Response to Orthostatic Stress*, PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, September 2004.
[11] I. C. F. Ipsen, C. T. Kelley, and S. R. Pope, *Nonlinear least squares problems and subset selection*. preprint, 2009.
[12] C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, no. 16 in Frontiers in Applied Mathematics, SIAM, Philadelphia, 1995.
[13] ———, *Iterative Methods for Optimization*, no. 18 in Frontiers in Applied Mathematics, SIAM, Philadelphia, 1999.
[14] K. Levenberg, *A method for the solution of certain nonlinear problems in least squares*, Quart. Appl. Math., 4 (1944), pp. 164–168.
[15] D. W. Marquardt, *An algorithm for least squares estimation of nonlinear parameters*, SIAM J., 11 (1963), pp. 431–441.
[16] S. Pope, *Parameter Identification in Lumped Compartment Cardiorespiratory Models*, PhD thesis, North Carolina State University, Raleigh, North Carolina, 2009.
[17] S. R. Pope, L. M. Ellwein, C. I. Zapata, V. Novak, C. T. Kelley, and M. S. Olufsen, *Estimation and identification of parameters in a lumped cerebrovascular model*, Math. Biosci. Eng., 6 (2008), pp. 93–115.
[18] K. Tanabe, *Continuous Newton-Raphson method for solving an underdetermined system of nonlinear equations*, Nonlinear Analysis, TMA, 3 (1979), pp. 493–501.
[19] ———, *Global analysis of continous analogs of the Levenberg-Marquardt and Newton-Raphson methods for solving nonlinear equations*, Annals of the Institute of Statistical Mathematics, 37 (1985), pp. 189–203.