Rank-Deficient Nonlinear Least Squares Problems and Subset Selection

Ilse Ipsen

North Carolina State University
Raleigh, NC, USA

Joint work with: Tim Kelley and Scott Pope

Overview

Motivating Application

Modeling cardiovascular systems
Extract biomarkers: Nonlinear parameter estimation
Nonlinear dependencies among parameters

Computation

Solution of nonlinear least squares problem
by Levenberg-Marquardt trust region algorithm
Rank deficient Jacobians
Errors in Jacobian evaluation

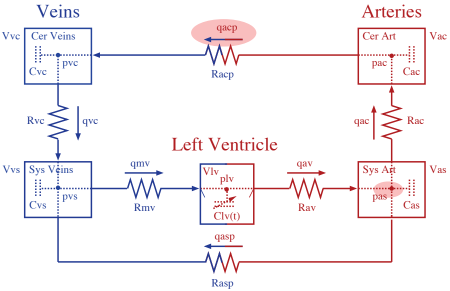
How to "regularize" the Jacobian?

Truncated SVD: NO
Column subset selection: YES

Modeling Cardiovascular Systems

Goal: Identify parameters that regulate blood flow

Cardiovascular system = lumped 5-compartment model
Blood flow, volume, pressure, resistance, compliance



[Pope, Olufsen, Ellwein, Novak, Kelley]

Computation

- System of 5 ODEs with $N = 16$ parameters

$$y' = F(t, y; p) \qquad y(0) = y_0$$

  Parameter vector $p \in R^N$

- Observations d_j at M time points t_j $\qquad M \gg N$
- Nonlinear residual

$$R(p) = \begin{pmatrix} y(t_1, p) - d_1 \\ \vdots \\ y(t_M, p) - d_M \end{pmatrix}$$

- Identify parameters p that minimize difference between measured and computed quantities

$$\min_p R(p)^T R(p)/2$$

Nonlinear Least Squares Problem

$$\min_p R(p)^T R(p)/2$$

Jacobian $J_n \equiv R'(p_n)$ at current iterate p_n

- Levenberg-Marquardt trust region algorithm

$$\texttt{For}\ n = 0, 1, 2 \ldots$$

$$p_{n+1} = p_n - \left(\nu_n I + J_n^T J_n\right)^{-1} J_n^T R(p_n)$$

- $\nu_n = 0$ and J_n full column rank: Gauss Newton
- Here: $\nu_n \geq 0$ and J_n rank deficient

Levenberg-Marquardt Algorithm

- Inside a Levenberg-Marquardt Step:

  `While` *iterate has not changed*

  Trial step $s = -\left(\nu_n I + J_n^T J_n\right)^{-1} J_n^T R(p_n)$

  Trial iterate $p_t = p_n + s$

  `if` p_t *good enough* `then`

  $p_{n+1} = p_t$, $\nu_{n+1} \leftarrow$ *keep or decrease* ν_n

  `else` $\nu_{n+1} \leftarrow$ *increase* ν_n

- Ideally:

  $\nu_n \rightarrow 0$, *or at least* ν_n *bounded*

  p_n *converge to minimizer, or at least stationary point*

- But here:

  Poor convergence (Levenberg-Marquardt stagnates)

  Gradient at "solution" not small

  Accuracy of "solution" ???

Convergence Analysis

- Near solution manifold
- Assuming exact arithmetic

Nonlinear iterations with rank deficient Jacobians:
Ben-Israel 1966, Boggs 1976, Deuflhard & Heindl 1979, Schaback 1985

Behavior of Levenberg-Marquardt Iterates

Assumptions

- Initial iterate p_0 close enough to a solution p^*
- $J(p)$ Lipschitz continuous
- $R(p^*)$ small but not necessarily zero

Then we can show

- Levenberg-Marquardt parameters ν_n remain bounded
- Iterates p_n approach solution manifold
- If p_n converge then they converge to some solution (Cauchy sequence)

Still need to show that p_n converge

Convergence of Levenberg-Marquardt Iterates

Model nonlinear dependence among parameters:

$$R(p) = \tilde{R}\left(B(p)\right) \qquad B : \mathbb{R}^M \to \mathbb{R}^K$$

If $K = N$ then Jacobian J has full column rank

Assumptions: Sufficiently close to a solution p^*

- $\tilde{R}$ and B uniformly Lipschitz continuously differentiable
- All K singular values of B' uniformly bounded away from 0
- All K singular values of $\tilde{R}'$ uniformly bounded away from 0
- $\tilde{R}(b)^T \tilde{R}(b)/2$ has unique minimizer

Then: p_n converge to a solution r-linearly

Summary: Convergence Analysis

Assumptions:

- Near solution manifold
  Initial iterate p_0 sufficiently close to a solution p^*
- Nonlinear residual $R(p^*)$ small but not necessarily zero
- Dependence among parameters:

$$R(p) = \tilde{R}\left(B(p)\right) \qquad B: \mathbb{R}^M \to \mathbb{R}^K$$

- Jacobians B' and $\tilde{R}'$ have rank K
- All Jacobians sufficiently smooth

Then: Iterates p_n converge to some solution

But this assumes exact arithmetic!
What happens in finite precision?

Finite Precision Issues

- Computation of trial step
- Effect of errors in computed Jacobian
- Regularization of Jacobian:
  Truncated SVD $\leftrightarrow$ subset selection

Singular vector perturbations: Stewart 1973

Computation of Trial Step

$$p_{n+1} = p_n + s$$

- Trial step

$$s = -\left(\nu I + J^T J\right)^{\dagger} J^T R$$

Works for $\nu = 0$ and rank deficient J

- Computed as minimum norm solution to linear least squares problem $\min_x \|Ax - b\|$

$$A = \begin{pmatrix} J \\ \sqrt{\nu} I \end{pmatrix} \qquad b = \begin{pmatrix} R \\ 0 \end{pmatrix}$$

Illconditioned or illposed if ν small, J rank deficient

Full Rank Jacobian

- J has full column rank: $\quad s = -\left(\nu I + J^T J\right)^{-1} J^T\, R$

  Condition number $\quad \kappa_\nu(J) := \left\| \left(\nu I + J^T J\right)^{-1} J^T \right\| \, \|J\|$

- $\tilde{J} = J + E$ has full column rank, $\quad \|E\| \leq \epsilon \|J\|$

  $\tilde{s} = -\left(\nu I + \tilde{J}^T \tilde{J}\right)^{-1} \tilde{J}^T\, R$

- Relative error in trial step

$$\frac{\|\tilde{s} - s\|}{\|\tilde{s}\|} \leq \kappa_\nu(J) \left(1 + \frac{\|R\|}{\|J\| \, \|\tilde{s}\|}\right) \epsilon$$

Error in J amplified by conditioning of J and nonlinear residual R

Rank Deficient Jacobian: Regularization

- Exact trial step $s = -\left(\nu I + J^T J\right)^{\dagger} J^T R$
- Truncate SVD of J: Truncated Jacobian J_t has singular values

$$\sigma_1 \geq \cdots \geq \sigma_K > \sigma_{K+1} = \cdots = \sigma_N = 0$$

- "Truncated" trial step is

$$s_t = -\left(\nu I + J_t^T J_t\right)^{\dagger} J_t^T R$$

- Computed as minimum norm solution of $\min_x \|A_t x - b\|$

$$A_t = \begin{pmatrix} J_t \\ \sqrt{\nu} I \end{pmatrix} \quad b = \begin{pmatrix} R \\ 0 \end{pmatrix}$$

Linear least squares problem still ill-conditioned or ill-posed

Truncated SVD: Errors in Jacobian

- SVD of exact Jacobian: $J = U\Sigma V^T \quad \mathrm{rank}(J) = K$

  Nonzero singular values $\sigma_1 \geq \cdots \geq \sigma_K > 0$

$$\kappa_\nu(J) := \sigma_1 \max_{\sigma_K \leq \sigma \leq \sigma_1} \frac{\sigma}{\nu + \sigma^2}$$

- SVD of perturbed Jacobian: $\quad J + E = \tilde{U}\tilde{\Sigma}\tilde{V}^T$, $\|E\|_F \leq \epsilon \|J\|$

  $\tilde{U}$, $\tilde{V}$ rotations of U, V by angles $\leq \theta$

- $\tilde{J}_t$ truncated SVD of $J + E \quad \mathrm{rank}(\tilde{J}_t) = K$

- Trial step of truncated perturbed Jacobian

$$\tilde{s}_t = -\left(\nu I + \tilde{J}_t^T \tilde{J}_t\right)^\dagger \tilde{J}_t^T R$$

Relative Error for Truncated SVD

For ϵ sufficiently small

$$\frac{\|\tilde{s}_t - s\|}{\|\tilde{s}_t\|} \leq \kappa_\nu(J) \left[1 + (1 + 2\|J\| \tan\theta) \frac{\|R\|}{\|J\| \, \|\tilde{s}_t\|}\right] \epsilon + \mathcal{O}(\epsilon^2)$$

- $\tan\theta$: Accuracy of singular vectors of truncated Jacobian $\tilde{J}_t$
- Error in J_t amplified by

  Conditioning of J
  Inaccuracy of singular vectors
  Nonlinear residual R

- Trial step from truncated SVD not accurate if

  J close to matrix of rank $K-1$: $\kappa_\nu(J) \gg 1$
  Singular vectors have low accuracy: $\tan\theta \gg 0$
  Nonlinear residual R large

Alternative Regularization: Subset Selection

- Choose K "very" linearly independent columns J_1 from J

$$J = \begin{pmatrix} J_1 & J_2 \end{pmatrix}$$

- Trial step

$$\hat{s} = -\left(\nu I + J_1^T J_1\right)^{-1} J_1^T R$$

- Computed as solution of $\min_x \|A_1 x - b\|$

$$A_1 = \begin{pmatrix} J_1 \\ \sqrt{\nu} I \end{pmatrix} \qquad b = \begin{pmatrix} R \\ 0 \end{pmatrix}$$

Linear least squares problem now well-conditioned

Subset Selection: Errors in Jacobian

- J_1 selected by strong RRQR [Gu & Eisenstat 1996]

$$\frac{\sigma_j}{\sqrt{1+K(N-K)}} \leq \sigma_j(J_1) \leq \sigma_j \qquad 1 \leq j \leq K$$

  Singular values of J_1 close to largest singular values of J

- Perturbed Jacobian $\tilde{J} = J + E$

  $\mathrm{rank}(J + E) \geq K, \quad \|E\| \leq \epsilon\|J\|$

- Select same columns for $\tilde{J}_1$ and J_1: $\quad \begin{pmatrix} \tilde{J}_1 & \tilde{J}_2 \end{pmatrix}$

$$\tilde{s} = -\left(\nu I + \tilde{J}_1^T \tilde{J}_1\right)^{-1} \tilde{J}_1^T R$$

Subset Selection: Relative Error

- Condition number

$$\tilde{\kappa}_\nu(J_1) = \sigma_1 \max_{\tilde{\sigma}_K \le \sigma \le \sigma_1} \frac{\sigma}{\nu + \sigma^2} \qquad \tilde{\sigma}_K = \sigma_K / \sqrt{1 + K(N-K)}$$

- Relative error in subset selection trial step

$$\frac{\|\tilde{s} - \hat{s}\|}{\|\tilde{s}\|} \le \tilde{\kappa}_\nu(J_1) \left(1 + \frac{\|R\|}{\|J\| \, \|\tilde{s}\|}\right) \epsilon$$

Error in J amplified by conditioning of J_1 and nonlinear residual R

Same as full rank bound applied to J_1

Numerical Experiments

Goal: Design simplest possible setting to reproduce failures from truncated SVD observed in cardiovascular model

Numerical Experiments

- Driven harmonic oscillator

$$(1 + 10^{-3}\,\delta)y'' + (c_1 + c_2)y' + k\,y = 2\sin(5t)$$

$y(0) = y_0$, $y'(0) = y_0'$

- 4 parameters $p = \begin{pmatrix} \delta & c_1 & c_2 & k \end{pmatrix}^T$
- Numerical solution $\tilde{y}(t_j)$ from Matlab `ode15s`
- Nonlinear residual

$$R(p) = \begin{pmatrix} \tilde{y}(t_1) - d_1 \\ \vdots \\ \tilde{y}(t_M) - d_M \end{pmatrix}$$

- Estimate p by solving nonlinear least squares problem

$$\min_p R(p)^T R(p)/2$$

Numerical Experiments: Assumptions

$$(1 + 10^{-3}\,\delta)y'' + (c_1 + c_2)y' + k\,y = 2\sin(5t)$$

- Highly accurate Jacobians:
  Compute columns of J from sensitivities $\partial y/\partial p$
- Zero residual:
  Data d from exact parameters $\quad p^* = \begin{pmatrix} 1.23 & 1 & 0 & 1 \end{pmatrix}^T$
- Initial guess $p_0 = \begin{pmatrix} 0 & 1 & 1 & .3 \end{pmatrix}^T$
- Singular values of initial Jacobian:

$$40.1 \quad 12.9 \quad 7.4 \cdot 10^{-4} \quad 6.21 \cdot 10^{-16}$$

- One zero singular value by design: $\quad \frac{\partial R}{\partial c_1} = \frac{\partial R}{\partial c_2}$
  Need to recover $c_1 + c_2 = 1$

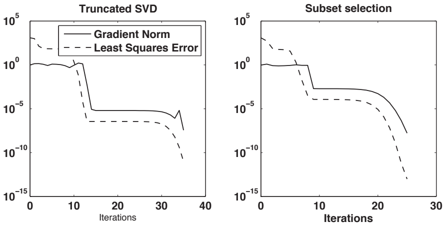
Numerical Experiments: Zero Residual

Assumptions for subset selection:

- Rank $K = 3$ of Jacobian is known
- Subset selection applied only to initial Jacobian
- All Levenberg-Marquardt iterations work with same K columns
- Parameters corresponding to $N - K = 1$ non-selected columns set to nominal values
- Exact parameters $\quad p^* = \begin{pmatrix} 1.23 & 1 & 0 & 1 \end{pmatrix}^T$

- Truncated SVD: $p = \begin{pmatrix} 1.22 & .5 & .5 & 1 \end{pmatrix}^T$
- Subset selection: $p = \begin{pmatrix} 1.23 & .5 & .5 & 1 \end{pmatrix}^T$
  A little more accurate

Convergence History: Zero Residual



Subset selection converges faster and slightly more accurate than truncated SVD

Numerical Experiments: Non-Zero Residual

$$(1 + 10^{-3}\delta)y'' + (c_1 + c_2)y' + k_0 y = 2\sin(5t)$$

- Non-zero residual
  Componentwise relative perturbation of data d by 10^{-4}
- Singular values of Jacobian have not changed:

$$40.1 \quad 12.9 \quad 7.4 \cdot 10^{-4} \quad 6.21 \cdot 10^{-16}$$

- Exact parameters $\quad p^* = \begin{pmatrix} 1.23 & 1 & 0 & 1 \end{pmatrix}^T$

- Truncated SVD: $p = \begin{pmatrix} .09 & .5 & .5 & .998 \end{pmatrix}^T$
  δ completely wrong
- Subset selection: $p = \begin{pmatrix} 1.28 & 0 & 1 & 1 \end{pmatrix}^T$
  Much more accurate

Truncated SVD $\longleftrightarrow$ Subset Selection

What is really going on?

General Least Squares Problems

$$\min_x \|Ax - b\| \qquad A \text{ is } M \times N, \quad M \geq N$$

Singular values $\quad \sigma_1 \geq \cdots \geq \sigma_K \gg \sigma_{K+1} \geq \cdots \geq \sigma_N > 0$

Least squares problem with illconditioned matrix

- Truncated SVD
  Singular values $\quad \sigma_1 \geq \cdots \geq \sigma_K \gg \sigma_{K+1} = \cdots = \sigma_N = 0$
  Least squares problem now ill-posed

- Subset Selection: K columns of A selected by strong RRQR
  Singular values $\quad \sigma_1 \geq \cdots \geq \sigma_K / \sqrt{1+K(N-k)} \gg 0$
  Least squares problem with wellconditioned matrix

The Problem with Truncated SVD

- $\min_x \|Ax - b\|$

  A has singular values $\sigma_1 \geq \cdots \geq \sigma_k \geq \cdots \geq \sigma_r > 0$
  $s = A^\dagger b$ is minimal norm solution

- Truncated SVD: $\min_x \|A_t x - b\|$

  A_t has singular values $\sigma_1 \geq \cdots \geq \sigma_k$
  $s_t = A_t^\dagger b$ is minimal norm solution, residual $r_t = b - As_t$

- Relative error

$$\frac{\|s_t - s\|}{\|s_t\|} \leq \frac{\sigma_1}{\sigma_r} \frac{\|r_t\|}{\|A\| \, \|s_t\|}$$

Small residual does not imply that s_t accurate
Bound independent of how many singular values truncated

Summary

- Parameter estimation with nonlinear dependences
- Expressed as nonlinear least squares problem
- Solved by Levenberg-Marquardt trust region algorithm
- Rank deficient Jacobians
- Errors in Jacobian evaluation, non-zero residuals

- How to regularize Jacobian:

  Truncated SVD: *NO*

  Subset selection: *Yes*