Part VIIc: Nonlinear Least Squares

MA 580; Numerical Analysis I

C. T. Kelley
NC State University
tim.kelley@ncsu.edu
Version of November 4, 2016

NCSU, Fall 2016
Part VIIc: Nonlinear Least Squares
Objective: recover spring constant and damping by fitting data to a model.

Model equation:

\[ u'' + cu' + ku = 0; \quad u(0) = u_0, \quad u'(0) = 0, \]

for \( 0 \leq t \leq T \).

Solve model equation with

- Analytic solution from elementary odes
- Numerical IVP solver: \texttt{ode15s}
Optimization Problem

- Optimization variable $\mathbf{x} = (c, k)^T \in \mathbb{R}^2$.
- Sample displacement at $\{t_j\}_{j=1}^M$, where $t_j = (j - 1)T / (M - 1)$
- Observations at $\{t_j\}_{j=1}^M$, where $t_j = (j - 1)T / (M - 1)$
- Data is $\mathbf{u} \in \mathbb{R}^M$ where $u_j \approx u(t_j)$
- Output of model is $u(t : \mathbf{x})$

Objective function:

$$f(\mathbf{x}) = \frac{1}{2} \sum_{j=1}^{M} |u(t_j : \mathbf{x}) - u_j|^2 \equiv \mathbf{R}(\mathbf{x})^T \mathbf{R}(\mathbf{x}) / 2.$$
How is this like an equation?

We’re aiming to make the norm of $R(x)$, the deviation from the data, small. Here

$$r_i(x) = u(t_j : x) - u_j$$

So this is to nonlinear equations as linear least squares is to linear equations. We should be able to get a decent algorithm out of this.
Instead of
\[ F(x) = 0 \]
we now do
\[ \min R(x)^T R(x)/2 \equiv \min \frac{\|R(x)\|^2}{2} \]
where \( R : \mathbb{R}^N \rightarrow \mathbb{R}^M \) and \( M \geq N \).
Necessary Conditions

From now on $M \geq N$!

The Jacobian is the $M \times N$ matrix

$$R'(x)_{ij} = \frac{\partial r_i(x)}{\partial x_j}$$

Work out the math to get

$$\nabla f(x) = R'(x)^T R(x) \in \mathbb{R}^N.$$  

The first-order necessary conditions are

$$R'(x^*)^T R(x^*) = 0.$$
What do the first-order conditions mean?

- Natural extension of normal equations

\[(Ax - b)^T(Ax - b) = A^T Ax - A^T b = 0.\]

- Sort of an extension of \( F(x) = 0 \)

But the second order necessary condition (\( \nabla^2 f \) spd) is not part of the package!

\[ \nabla^2 f(x) = \ldots \]
Local Model

Approximate min $f$ by the linear least squares problem

$$\min \| R'(x_c)(x_+ - x_c) + R(x_c) \|^2 = 0$$

Properties of both nonlinear equations and optimization.
If $R'$ has full column rank (which we assume) then the step

$$s = x_+ - x_c = -(R'(x_c)^T R'(x_c))^{-1} R(x_c)$$

and

$$x_+ = x_c + s.$$
Evaluate $R(x_c)$ and $\|R'(x_c)^T R(x_c)\|$ Terminate?
Factor $R'(x_c) = QR$.
$s = -R^{-1}Q^T R(x_c)$
$x_+ = x_c + s$. 

Standard Assumptions $M \geq N$

- $R'(x^*)^T R(x^*) = 0$
- $R'$ has full column rank.
- $R'$ is Lipschitz continuous with Lipschitz constant $\gamma$.

We can replace the first assumption by $x^*$ is a local minimizer or (sorta) derive that.

Then (derive this)

$$R(x)^T R(x) = \|R(x^*)\|^2 + \|R'(x^*)e\|^2 + O(\|R(x^*)\|\|e\|^2)$$

and hence $x^*$ is a local minimizer if $\|R(x^*)\|$ is sufficiently small.
Theorem: SA + Good Data imply that

\[ \|e_+\| = O(\|e\|^2 + \|R(x^*)\| \|e\|) \]

So Gauss-Newton converges if \(\|R(x^*)\|\) and \(\|e_0\|\) are sufficiently small.
Proof: 1

Pick \( \delta > 0 \) such that \( \| e_c \| \leq \delta \) implies \( R' \) has full rank.

By definition of the Gauss-Newton step

\[
e_+ = e_c - (R'(x_c)^T R'(x_c))^{-1} R'(x_c)^T R(x_c)
\]

\[
= (R'(x_c)^T R'(x_c))^{-1} R'(x_c)^T (R'(x_c)e_c - R(x_c)).
\]

By Taylor's theorem

\[
R'(x_c)e_c - R(x_c) = R'(x_c)e_c - R(x^*) + R(x^*) - R(x_c)
\]

\[
= -R(x^*) + (R'(x_c)e_c + R(x^*) - R(x_c)).
\]
Plug this into the equation for $e_+$ to bet

\[ R'(x_c)^T R'(x_c) R'(x_c) e_+ = R'(x_c)^T (R'(x_c) e_c - R(x_c)) \]

\[ = -R'(x_c)^T R(x^*) + a, \]

where

\[ a = R'(x_c)^T (R'(x_c) e_c + R(x^*) - R(x_c)). \]
Proof: III

Now,

$$\|R'(x_c)e_c + R(x^*) - R(x_c)\| \leq \gamma \|e_c\|^2 / 2$$

so

$$\|a\| \leq \gamma \|R'(x_c)\|\|e_c\|^2 / 2$$

$R'(x^*)^T R(x^*) = 0$, so

$$-R'(x_c)^T R(x^*) = (R'(x^*) - R'(x_c))^T R(x^*).$$

which implies that . . .
Proof: IV

\[ \|e_+\| \leq \left\| (R'(x_c)^T R'(x_c))^{-1} \left( (R'(x^*) - R'(x_c))^T R(x^*) \right) \right\| 
\]

\[ + \left( \frac{(R'(x_c)^T R'(x_c))^{-1} \| R'(x_c)^T \gamma \| e_c \|^2}{2} \right) \]

\[ \leq \left( (R'(x_c)^T R'(x_c))^{-1} \| e_c \| \right) \left( \| R(x^*) \| + \| R'(x_c)^T \| \| e_c \| / 2 \right) \]

\[ = O(\| R(x^*) \| \| e_c \| + \| e_c \|^2). \]
What Gauss-Newton is not.

You are not solving

$$\nabla f(x) = R'(x)^T R(x) = 0$$

with Newton. That’s not a good idea because

- The Jacobian of $\nabla f$ has $R''$ in it.
- Solving $\nabla f(x) = 0$ does not imply minimization.
Example: Parameter ID problem

- Data: solution with $c = k = 1$, $u(0) = 10$, $u'(0) = 0$
  
  $$u(t) = 10e^{-t/2} \cos \left( \frac{\sqrt{3}t}{2} \right)$$

- Evaluate $R$ with ode15s with tolerances $10^{-8}$
  convert to first-order system . . .

- Two cases
  - Exact solution (zero residual)
  - Exact solution with .01% random error (non-zero residual)
What’s in the plots.

- Residual norm: $\|R(x)\|

  Remember $f(x) = \|R(x)\|^2/2$

- Gradient norm: $\|R'(x)^T R(x)\|

  Remember $\nabla f(x) = R'(x)^T R(x)$
Part VIIc: Nonlinear Least Squares

Results: zero residual, limited by solver accuracy

![Iteration history, zero residual](image)
Results: non-zero residual, limited by residual norm

Iteration history, non-zero residual

- Residual Norm
- Gradient Norm