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

Parameter Identification for Initial Value Problem

Objective: recover spring constant and damping by fitting data to a model.
Model equation:

$$u'' + cu' + ku = 0; u(0) = u_0, u'(0) = 0,$$

for $0 \leq t \leq T$.
Solve model equation with

- Analytic solution from elementary odes
- Numerical IVP solver: `ode15s`

Optimization Problem

- Optimization variable $\mathbf{x} = (c, k)^T \in R^2$.
- 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 $\mathbf{u} \in R^M$ 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 $\mathbf{R}(\mathbf{x})$, the deviation from the data, small. Here

$$r_i(\mathbf{x}) = u(t_j : \mathbf{x}) - u_j$$

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.

Nonlinear Least Squares

Instead of

$$\mathbf{F}(\mathbf{x}) = 0$$

we now do

$$\min \mathbf{R}(\mathbf{x})^T \mathbf{R}(\mathbf{x})/2 \equiv \min \frac{\|\mathbf{R}(\mathbf{x})\|_2^2}{2}$$

where $\mathbf{R} : R^N \rightarrow R^M$ and $M \geq N$.

Necessary Conditions

From now on $M \geq N$!
The Jacobian is the $M \times N$ matrix

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

Work out the math to get

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

The first-order necessary conditions are

$$\mathbf{R}'(\mathbf{x}^*)^T \mathbf{R}(x^*) = 0.$$

What do the first-order conditions mean?

- Natural extension of normal equations

$$(\mathbf{Ax} - \mathbf{b})^T(\mathbf{Ax} - \mathbf{b}) = \mathbf{A}^T\mathbf{Ax} - \mathbf{A}^T\mathbf{b} = 0.$$

- Sort of an extension of $\mathbf{F}(\mathbf{x}) = 0$

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

$$\nabla^2 f(\mathbf{x}) = \dots$$

Local Model

Approximate $\min f$ by the linear least squares problem

$$\min \|\mathbf{R}'(\mathbf{x}_c)(\mathbf{x}_+ - \mathbf{x}_c) + \mathbf{R}(x_c)\|^2 = 0$$

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

$$\mathbf{s} = \mathbf{x}_+ - \mathbf{x}_c = -(\mathbf{R}'(\mathbf{x}_c)^T \mathbf{R}'(\mathbf{x}_c))^{-1} \mathbf{R}(x_c)$$

and

$$\mathbf{x}_+ = \mathbf{x}_c + \mathbf{s}.$$

Algorithmic Description

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

Standard Assumptions $M \geq N$

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

We can replace the first assumption by $\mathbf{x}^*$ is a local minimizer or (sorta) derive that.
Then (derive this)

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

and hence $\mathbf{x}^*$ is a local minimizer if $\|\mathbf{R}(\mathbf{x}^*)\|$ is sufficiently small.

Convergence of Gauss-Newton

Theorem: SA + Good Data imply that

$$\|\mathbf{e}_+\| = O(\|\mathbf{e}\|^2 + \|\mathbf{R}(\mathbf{x}^*)\| \|\mathbf{e}\|)$$

So Gauss-Newton converges if $\|\mathbf{R}(\mathbf{x}^*)\|$ and $\|\mathbf{e}_0\|$ are sufficiently small.

Proof: I

Pick $\delta > 0$ such that $\|\mathbf{e}_c\| \leq \delta$ implies $\mathbf{R}'$ has full rank.
By definition of the Gauss-Newton step

$$\begin{aligned} \mathbf{e}_+ &= \mathbf{e}_c - (\mathbf{R}'(\mathbf{x}_c)^T \mathbf{R}'(\mathbf{x}_c))^{-1} \mathbf{R}'(\mathbf{x}_c)^T \mathbf{R}(\mathbf{x}_c) \\ &= (\mathbf{R}'(\mathbf{x}_c)^T \mathbf{R}'(\mathbf{x}_c))^{-1} \mathbf{R}'(\mathbf{x}_c)^T (\mathbf{R}'(\mathbf{x}_c)\mathbf{e}_c - \mathbf{R}(\mathbf{x}_c)). \end{aligned}$$

By Taylor's theorem

$$\begin{aligned} \mathbf{R}'(\mathbf{x}_c)\mathbf{e}_c - \mathbf{R}(\mathbf{x}_c) &= \mathbf{R}'(\mathbf{x}_c)\mathbf{e}_c - \mathbf{R}(\mathbf{x}^*) + \mathbf{R}(\mathbf{x}^*) - \mathbf{R}(\mathbf{x}_c) \\ &= -\mathbf{R}(\mathbf{x}^*) + (\mathbf{R}'(\mathbf{x}_c)\mathbf{e}_c + \mathbf{R}(\mathbf{x}^*) - \mathbf{R}(\mathbf{x}_c)). \end{aligned}$$

Proof: II

Plug this into the equation for $\mathbf{e}_+$ to bet

$$\begin{aligned}\mathbf{R}'(\mathbf{x}_c)^T\mathbf{R}'(\mathbf{x}_c)\mathbf{R}'(\mathbf{x}_c)\mathbf{e}_+ &= \mathbf{R}'(\mathbf{x}_c)^T(\mathbf{R}'(\mathbf{x}_c\mathbf{e}_c - \mathbf{R}(\mathbf{x}_c)\\ &= -\mathbf{R}'(\mathbf{x}_c)^T\mathbf{R}(\mathbf{x}^*) + \mathbf{a},\end{aligned}$$

where

$$\mathbf{a} = \mathbf{R}'(\mathbf{x}_c)^T(\mathbf{R}'(\mathbf{x}_c)\mathbf{e}_c + \mathbf{R}(\mathbf{x}^*) - \mathbf{R}(\mathbf{x}_c)).$$

Proof: III

Now,

$$\|\mathbf{R}'(\mathbf{x}_c)\mathbf{e}_c + \mathbf{R}(\mathbf{x}^*) - \mathbf{R}(\mathbf{x}_c)\| \le \gamma \|\mathbf{e}_c\|^2/2$$

so

$$\|\mathbf{a}\| \le \gamma \|\mathbf{R}'(\mathbf{x}_c)\| \|\mathbf{e}_c\|^2/2$$

$\mathbf{R}'(\mathbf{x}^*)^T \mathbf{R}(\mathbf{x}^*) = 0$, so

$$-\mathbf{R}'(\mathbf{x}_c)^T \mathbf{R}(\mathbf{x}^*) = (\mathbf{R}'(\mathbf{x}^*) - \mathbf{R}'(\mathbf{x}_c))^T \mathbf{R}(\mathbf{x}^*).$$

which implies that ...

Proof: IV

$$
\begin{aligned}
\|\mathbf{e}_+\| &\le \|(\mathbf{R}'(\mathbf{x}_c)^T\mathbf{R}'(\mathbf{x}_c))^{-1}\|\|(\mathbf{R}'(\mathbf{x}^*) - \mathbf{R}'(\mathbf{x}_c))^T\mathbf{R}(\mathbf{x}^*)\| \\
&+ \frac{\|(\mathbf{R}'(\mathbf{x}_c)^T\mathbf{R}'(\mathbf{x}_c))^{-1}\|\|\mathbf{R}'(\mathbf{x}_c)^T\|\gamma\|\mathbf{e}_c\|^2}{2} \\
&\le \|(\mathbf{R}'(\mathbf{x}_c)^T\mathbf{R}'(\mathbf{x}_c))^{-1}\|\gamma\|\mathbf{e}_c\|\left(\|\mathbf{R}(\mathbf{x}^*)\| + \|\mathbf{R}'(\mathbf{x}_c)^T\|\|\mathbf{e}_c\|/2\right) \\
&= O(\|\mathbf{R}(\mathbf{x}^*)\|\|\mathbf{e}_c\| + \|\mathbf{e}_c\|^2).
\end{aligned}
$$

What Gauss-Newton is not.

You are not solving

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

with Newton. That's not a good idea because

- The Jacobian of ∇f has $\mathbf{R}''$ in it.
- Solving $\nabla f(\mathbf{x}) = 0$ does not imply minimization.

Example: Parameter ID problem

- Data: solution with $c = k = 1$, $u(0) = 10$, $u'(0) = 0$

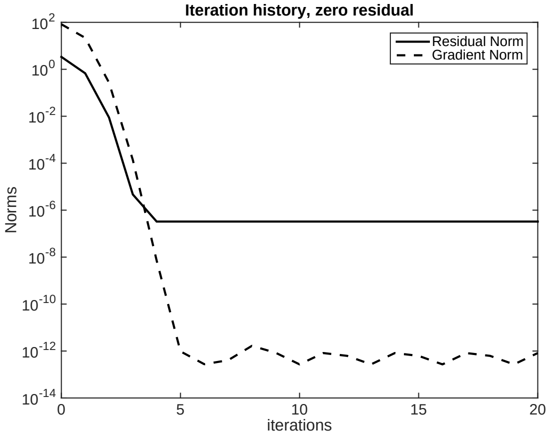
$$u(t) = 10e^{-t/2} \cos\left(\frac{\sqrt{3}t}{2}\right)$$

- 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: $\|\mathbf{R}(\mathbf{x})\|$
  Remember $f(\mathbf{x}) = \|\mathbf{R}(\mathbf{x})\|^2/2$
- Gradient norm: $\|\mathbf{R}'(\mathbf{x})^T\mathbf{R}(\mathbf{x})\|$
  Remember $\nabla f(\mathbf{x}) = \mathbf{R}'(\mathbf{x})^T\mathbf{R}(\mathbf{x})$

Results: zero residual, limited by solver accuracy



Results: non-zero residual, limited by residual norm