Model Reduction for Nonlinear Least Squares

C. T. Kelley

Joint work with Dan Sorensen, Jill Reese, Corey Winton

Department of Mathematics

Center for Research in Scientific Computation

North Carolina State University

Raleigh, North Carolina, USA

Workshop on PDE-Constrained Optimization

Oberwolfact, February 2006

Supported by NSF, ARO, DOEd.

Outline

- Model problem: Calibration of groundwater flow model
  - Surrogate models vs reduced model
  - Construction of reduced model
- Optimization via Pseudo-Transient Continuation (Ψtc)
- 1-D example
- 2-D example

Model Problem

Darcy's law for groundwater flow says:

$$div(K\nabla u) = f$$

where K is the spatially dependent hydraulic conductivity. Our objective is to approximate K from sparse measurements.

Standard approach

Banks/Kunisch 89, Doherty (PEST) 90's – present

- Parameterize K (spline, piecewise constant ...) by $p \in R^N$.
- Organize measurments into data vector $d \in R^M$.
- Write solver for discrete PDE to obtain solutions $u \in R^{M_x}$ when given p.
- Map u to data space with $D : R^{M_x} \longrightarrow R^M$ evaluation at well locations, for example.
- Solve $\min \|D(u(p)) - d\|_2^2$ or a regularized version of that problem.

For us: $M << N << M_x$, so the PDE solve is the expensive part

Surrogate Models

- Replace $\min f$ by $\min \bar{f}$ where $\bar{f}$ is inexpensive
  - Response surface:
    quadratic, radial basis, neural net, . . .
  - Coarse mesh version of PDE:
    different grid, less physics, . . .
  - Model reduction: Original PDE + smaller basis
    Captures problem structure (still least squares)
    Same code and same physics

Bulding the reduced model

- Discretize PDE with $A(p)u = f$.
- Find basis $\bar{U} = [u_1, \ldots, u_K]$ that "captures" most solutions.
- Replace $Au = f$ with

$$\bar{A}\bar{u} = \bar{U}^T A \bar{U}\bar{u} = \bar{f} = \bar{U}^T f$$

So how do you get U?

PODS

Proper Ortogonal Decomposition from fluid control (Karhunen, 46)

- Collect snapshots $W = [w_1, w_2, \dots w_L]$ from time dependent simulation.
- Take SVD of snapshots: $U \Sigma V^T = W$.
- Identify K for which σ_{K+1} is “small”.
- $\bar{U} = [u_1, \dots, u_K]$

What’s L? What does “small” mean?

Artificial time dependent problem

- Write problem as $\min f$ where $f = R^T R/2$.
- $\nabla f(p) = R'(p)^R(p)$
- Integrate $p' = -\nabla f(p)$ for a few Euler steps. Collect the $u's$ to get W.
- Proceed as in POD

Optimization via Ψtc

Pseudo-Transient Continuation finds steady state solutions of

$$\frac{du}{dt} = -F(u)$$

by mimicing integration to steady state **with the goal of increasing the time step**.
Simple forumlation

$$u_{n+1} = u_n - (\delta_n^{-1} I + F'(u_n))^{-1} F(u_n)$$

where $\{\delta_n\}$ is controled by Switched Evolution Relaxation:

(A)$\delta_{n+1} = \delta_n / \|F(u_n)\|$ or (B)$\delta_{n+1} = \delta_n / \|x_n - x_{n-1}\|$

Optimization

General Idea: Higham, 1999 (also Fletcher 1987)

- $\min f \to u' = -\nabla f$ (very old idea)
- Solve with Ψtc , manage step with TR approach

Liao-Qi-Qi 2004, Liao-Qi-K 2006

- Constraints $\to$ nonsmooth gradient
- Use generalized derivative and/or smoothing
- Ψtc with SER/TR step control

Least Squares Example

Problem:

$$\min f(x) \text{ where } f(x) = R^T(x)R(x)/2,$$

$R : R^N \to R^M$, $M > N$.

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

Gauss-Newton approximation to $\nabla^2 f$ is $H(x) = R'(x)^T R'(x)$

Ψtc for nonlinear least squares

$$x_{n+1} = x_n - (\delta_n^{-1} I + R'(x_n)^T R(x_n))^{-1} R'(x_n)^T R(x_n)$$

Levenberg-Marquardt if we use no second derivative information.
Differences: management of δ (but see K. 1999)

Bound Constrained Problems

Problem: $\min_{x \in \Omega} f(x)$ where

$$\Omega = \{x \,|\, L_i \leq (x)_i \leq U_i\}$$

Necessary conditions for optimality

$$F(x) = x - \mathscr{P}(x - \nabla f(x)) = 0$$

where

$$\mathscr{P}(x)_i = \begin{cases} L_i & \text{if } (x)_i \leq L_i \\ (x)_i & \text{if } L_i < (x)_i < U_i \\ U_i & \text{if } (x)_i \geq U_i \end{cases}$$

Ψtc for bound constraints

The dynamics

$$x' = -F(x)$$

are stable and $\liminf \|F(x)\| = 0$ (Liao-Qi-Qi)
But F is nonsmooth, in a direct Ψtc

$$x_{n+1} = x_n - (\delta_n^{-1} I + F'(x_n))^{-1} F(x_n)$$

you have to approximate F' carefully. If you do this, convergence results of (K, Fowler 06) hold.
There's an easier way.

Projected Ψtc

$$x_{n+1} = \mathscr{P}(x_n - (\delta_n^{-1} I + H_n^r)^{-1} F(x_n))$$

where H_n^r is the reduced Hessian.
Contrast with scaled gradient projection

$$x_{n+1} = \mathscr{P}(x_n - (H_n^r)^{-1} \nabla f(x_n)).$$

Reduced Model Hessian

Given x_n, H_n, ε_n, let D_n be the diagonal matrix

$$(D_n)_{ii} = \begin{cases} 1 & \text{if } \|u_n - \mathscr{P}(u_n)\| > \varepsilon_n \\ 0 & \text{otherwise} \end{cases}$$

$$H_n^r = I - D_n(I - H_n)D_n$$

Three versions

Direct Ψtc :

$$x_{n+1} = x_n - (\delta_n^{-1} I + F'(x_n))^{-1} F(x_n)$$

Projected Ψtc :

$$x_{n+1} = \mathscr{P}(x_n - (\delta_n^{-1} I + H_n^r)^{-1} F(x_n))$$

Projected gradient projection:

$$x_{n+1} = \mathscr{P}(x_n - (H_n^r)^{-1} \nabla f(x_n)).$$

Manage δ with SER or Trust Region

Convergence?

Global and locally fast convergence if:

- Direct Ψtc : $x(t) \to x^*$; SER or TR δ management
- Projected GP: H_n^r uniformly well conditioned + spd
- Projected Ψtc : $x(t) \to x^*$
  H_n^r either spd (TR) or inexact Newton condition (SER)

Our experiments (Liao, K, 06; this talk) say that SER works best.
SER(B) is best for this application.