UNEDF Workshop
Park Forest, Washington
June 23 – 26, 2008

Broyden's Method for Large Systems of Nonlinear Equations

Jorge Moré

Joint work with Jason Sarich

Mathematics and Computer Science Division
Argonne National Laboratory

Argonne NATIONAL LABORATORY

Optimization Applications in SciDAC

- Quantum Chemistry
  - Energy minimization
  - Transition states
- Nuclear Physics
  - Nonlinear eigenvalues
  - Parameter estimation
  - Least action pathways
- Accelerator Design
  - Shape optimization
  - Nonlinear eigenvalues
- Groundwater Flow
  - Parameter estimation

TOPS

Towards Optimal Terascale Simulations

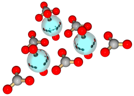
Have optimization technology,
will tackle applications

Toolkit for Advanced Optimization (TAO)

An optimization toolkit for solving large-scale optimization problems on advanced (massively parallel) architectures.

- ◇ Portability, performance, scalability
- ◇ An interface independent of architecture
- ◇ Leverage existing parallel computing infrastructure (PETSc)

TAO (www.mcs.anl.gov/tao)



NWChem, MPQC

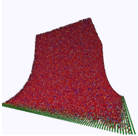
- ◇ Source code and documentation
- ◇ Installation instructions, example problems, . . .

Argonne NATIONAL LABORATORY

TAO Impact

Selected applications

- Semiconductor modelling
- Magnetic nanostructures
- Subsurface remediation
- Variational surfaces

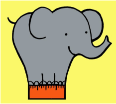


P. Bauman

P. Joshi

Toolkits

- `TADM` - Parameter estimation
- `BUSTER` - Protein structures
- `ELEFANT` - Statistical machine learning



UNEDF: Parameter Estimation in Nuclear Fission

$$f(x) = \sum_{k=1}^{m} \sigma_k \| f_k(x) - y_k \|^2 \qquad x \Longrightarrow \boxed{\text{HFODD}} \Longrightarrow f_k(x)$$

- Expensive evaluation of $f_k(x)$ ($U_{236} \approx$ 12 hours)
- Large memory requirements ($U_{236} \approx$ 0.5GB)
- Many nuclei (about 2,000)
- A wide range of observables (binding energy, . . .)
- Noisy function evaluations
- Lack of derivatives with respect to parameters
- Several minima with different predictive powers

UNEDF Research Issues: Nonlinear Optimization

- What are the best techniques for solving nonlinear, noisy optimization problems

$$\min \{ f(x) : x_L \le x \le x_U \}$$

when the gradient ∇f of f is not available and the evaluation of f is computationally intensive (1,000 CPU days)

- How can we solve systems of n nonlinear equations

$$H(x) = 0$$

when derivatives are not available and the number of variables n is large?

Broyden's Method: A Biased Bibliography

- Broyden (1965). Introduced two methods. *...since Method 2 apears to be unsatisfactory in practice ...*
- Gay and Schnabel (1977). Projected updates.
- Gay (1979). Broyden's method for linear systems. Refers to the second method as Broyden's **bad** method.
- Srivasta (1984). Modified Broyden's second method so that only a few vectors of storage of order n are needed.
- Johnson (1988). Modified Srivasta's approach to incorporate additional information from previous iterates.
- Byrd, Nocedal, Schnabel (1994). Compact form.
- Baran, Bulgac, Forbes, Hagen, Nazarewicz, Schunk, Stoitsov (2008). Improvement of Johnson's approach

Systems of Nonlinear Equations

Given a function $f : \mathbb{R}^n \to \mathbb{R}^n$, find a vector x^* that solves the system of nonlinear equations

$$f(x) = 0$$

Notation

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \qquad f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix}$$

The Jacobian matrix is

$$f'(x) = (\partial_1 f(x), \ldots, \partial_n f(x))$$

Broyden's Method 1

Given a sequence of iterates $x_0, x_1, \ldots$, Broyden's method generates approximation to the Jacobian matrix $f'(x_k)$ via

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k) s_k^T}{\|s_k\|^2},$$

where the vectors y_k and s_k are defined by

$$y_k = f(x_k + s_k) - f(x_k), \qquad s_k = x_{k+1} - x_k$$

Note that B_{k+1} satisfies the secant equations

$$B_{k+1} s_k = y_k$$

Broyden's Method 2

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k) y_k^T}{\|y_k\|^2}$$

Compact Form

$$H_{k+1} = H_0 + (S_k - H_0 Y_k) N_k^{-1} Y_k^T$$

where

$$[N_k]_{i,j} = \begin{cases} y_i^T y_j & \text{if } i \le j \\ 0 & \text{otherwise} \end{cases}$$

$$S_k = [s_1, ..., s_k], \qquad Y = [y_1, ..., y_k]$$

Broyden's Method 2: Projected Updates

$$H_{k+1} = H_0 + (S_k - H_0 Y_k)\left(Y_k^T Y_k\right)^{-1} Y_k^T$$

where

$$S_k = [s_1, ..., s_k], \qquad Y = [y_1, ..., y_k]$$

Remarks The projection into the space spanned by Y_k is

$$P_k = \left(Y_k^T Y_k\right)^{-1} Y_k^T$$

Note that

$$H_{k+1} Y_k = S_k$$

and thus secant equations from previous iterations are satisfied.

Johnson (1988) method

In our notation, Johnson's method is of the form

$$x_{k+1} = x_k - J_k f(x_k),$$

where J_k depends on constants α, w_0, and w_j for $j = 1, \ldots m$.

Theorem. If $H_0 = -\alpha I$, $w_0 = 0$, and $w_j \equiv 1$, then Johnson's method is the projected Broyden method, that is,

$$J_k = H_0 + (S_k - H_0 Y_k)\left(Y_k^T Y_k\right)^{-1} Y_k^T$$

Implementation for Large Problems: The Main Issues

- Broyden's method 1 or Broyden's method 2?
- How do we globalize Broyden? A line search method

$$x_{k+1} = x_k - \alpha_k H_k f(x_k)?$$

  But the search direction $-H_k f(x_k)$ may not be downhill.
- At each iteration we keep the last $s_1, \ldots, s_m$ and $y_1, \ldots, y_m$. How many vectors of storage should we keep?
- Do we need to re-start? At some point it may be better to throw away all previous information and re-start with the current best iterate x_k and some inverse Hessian H_k.

Computational Experiments: Benchmark Problems

Eight nonlinear variational problems of the form

$$\min \{ V(x) : x \in \mathbb{R}^n \}$$

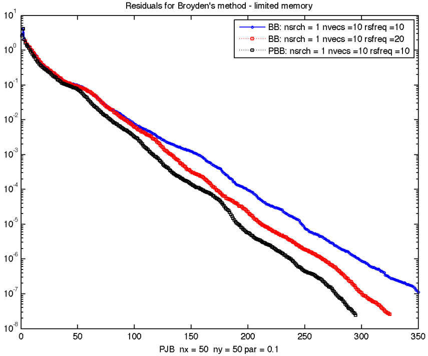
The nonlinear systems are the gradient equations

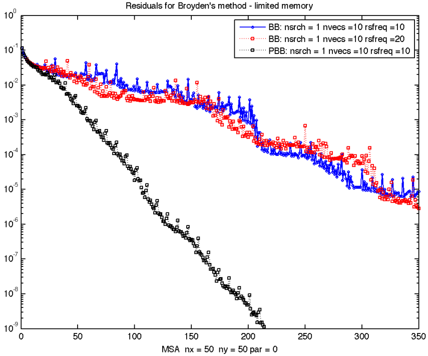
$$f(x) = \nabla V(x)$$

We show results for 4 problems with $n = 2500$ variables.

- PJB. Pressure in a journal bearing. Convex, ill-conditioned
- MSA. Minimal surface area. Convex, ill-conditioned
- ODC. Optimal design. Convex, discontinuos Jacobian $\nabla^2 f$
- GL2. Ginzburg-Landau conductor. Non-convex, ill-conditioned

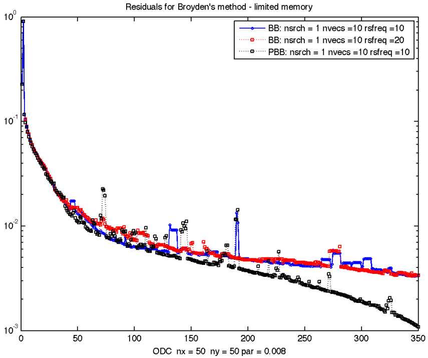
Computational Experiments: BB or PBB

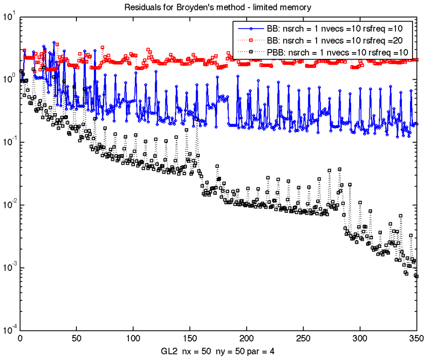




Argonne NATIONAL LABORATORY

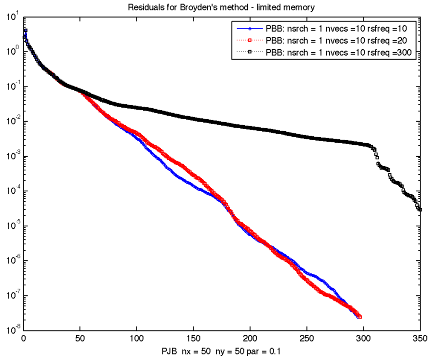
Computational Experiments: BB or PBB

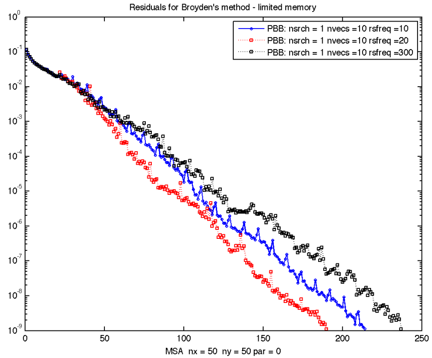




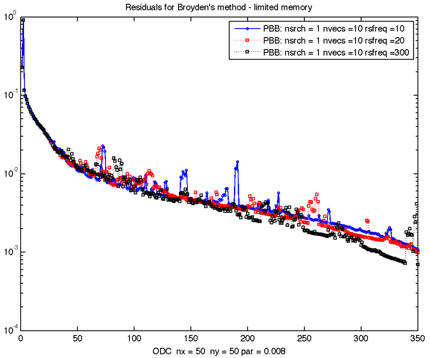
Argonne NATIONAL LABORATORY

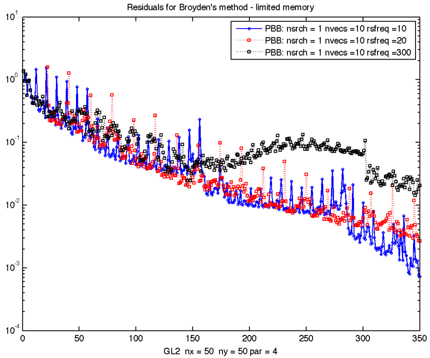
Computational Experiments: Restart Frequency





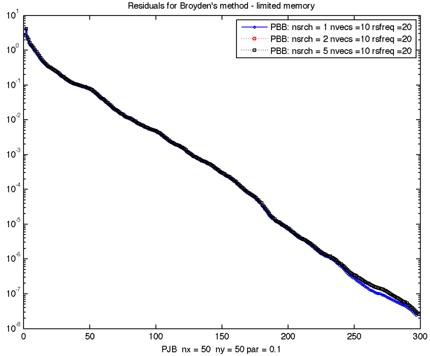
Computational Experiments: Restart Frequency

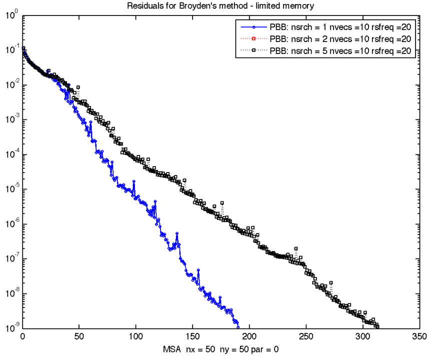




Argonne NATIONAL LABORATORY

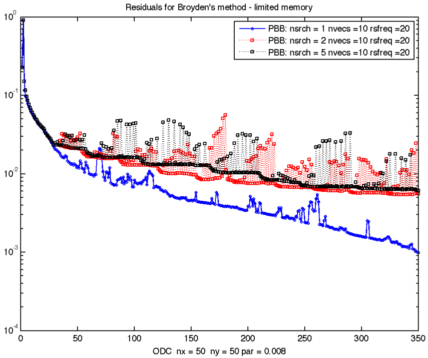
Computational Experiments: Number of Trial Points

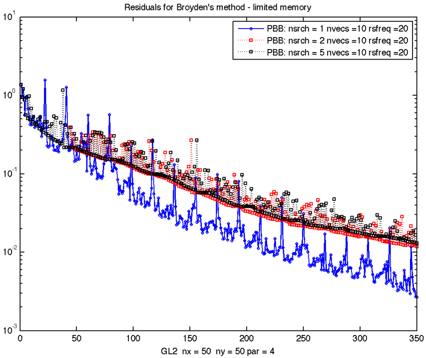




Argonne NATIONAL LABORATORY

Computational Experiments: Number of Trial Points





Argonne NATIONAL LABORATORY

Contributions: Optimization and Performance Analysis

- HFODD majordomo list (January 2007)
- Profile of HFODD using Tuning and Analysis Utilities (TAU)
- HFODD re-structuring (40% computing time reduction)
- NEDFT planning meeting (March 2007)
- Full storage eigenvalue solver (60% computing time reduction)
- NEDFT planning workshop (August 2007)
- Analysis of Broyden's method
- Nuclear Energy Functional workshop (January 2008)
- Optimization in SciDAC Applications, J. of Physics (2007)
- Benchmarking derivative-free optimization algorithms (2007)

Future Work

Year 3

- Development and performance of mass-table algorithms in BG
- Parallelization of HFODD
- Development of model-based derivative-free algorithms

Year 4

- Model and geometry-based optimization algorithms for DFT
- Investigation of performance on new DFT functionals

Year 5

- Fission pathways
- Performance, evaluation, and validation of DFT functional

Argonne NATIONAL LABORATORY