Second LACM-EFES-JUSTIPEN Workshop
Oak Ridge, Tennessee
January 23 – 25, 2008

A Short Guide to Optimization Technology on High-Performance Architectures

Jorge Moré

Joint work with Todd Munson and Jason Sarich

Mathematics and Computer Science Division
Argonne National Laboratory

Argonne NATIONAL LABORATORY

TOPS 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

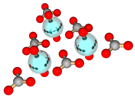
Argonne NATIONAL LABORATORY

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)



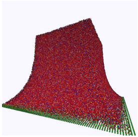
NWChem, MPQC

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

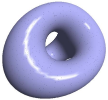
TAO Impact

Selected applications

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



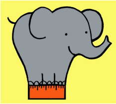
P. Bauman



P. Joshi

Toolkits

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

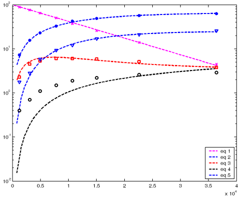


Noisy Optimization Problems: α-pinene

Determine the reaction coefficients in the thermal isomerization of α-pinene from measurements $z_1, \ldots z_8$ by minimizing

$$\sum_{j=1}^{8} \|y(\tau_j;\theta) - z_j\|^2$$

$$\begin{aligned}
y_1' &= -(\theta_1 + \theta_2) y_1 \\
y_2' &= \theta_1 y_1 \\
y_3' &= \theta_2 y_1 - (\theta_3 + \theta_4) y_3 + \theta_5 y_5 \\
y_4' &= \theta_3 y_3 \\
y_5' &= \theta_4 y_3 - \theta_5 y_5
\end{aligned}$$



This problem can be formulated as a (noisy) optimization problem

$$\min \{ f(\theta) : \theta \geq 0 \}$$

Argonne NATIONAL LABORATORY

Noisy Bound-Constrained Optimization

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

What are the best optimization techniques for nonlinear, noisy (computationally-intensive) problems?

Approaches

- Stochastic methods (simulated annealing, . . .)
- Geometry-based methods (pattern search, . . .)
- Model-based methods (quadratic models, . . .)
- Gradient methods (conjugate gradient, . . .)

Derivative-Free Optimization

Geometry-based Methods

- Nelder-Mead
- Pattern search

Pattern $\mathcal{P}_k$

$$\mathcal{P}_k \Longrightarrow \mathcal{P}_{k+1}$$

Model-based Methods

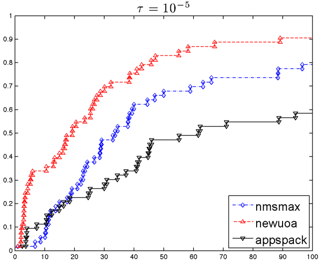
- Quadratic models
- Radial-basis models

Model q_k

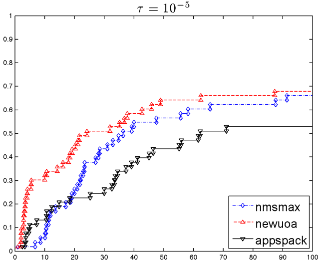
$$q_k \Longrightarrow q_{k+1}$$

Numerical Recipes only mentions the Nelder-Mead method

Performance of Derivative-Free Methods



Smooth problems



Noisy problems $\left(\varepsilon_F = 10^{-3}\right)$

- Data profiles with $f(x) \leq f(x^*) + \tau(f(x_0) - f(x^*))$
- Units are the number of (simplex) gradients for convergence
- Benchmark problems with n variables, $2 \leq n \leq 16$

Reference: www.mcs.anl.gov/~more/dfo

Argonne NATIONAL LABORATORY

Gradient-Based Optimization Methods

Assume that the gradient ∇f is available. What are the best techniques for the optimization problem

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

Nonlinear conjugate gradient methods

- Fletcher-Reeves (FR)
- Polak-Rivière (PR)
- Limited-memory CG methods

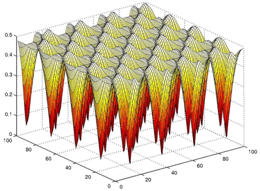
Numerical Recipes only mentions FR and PR.

TAO Benchmark: Ginzburg-Landau Model

Minimize the Gibbs free energy for a homogeneous superconductor

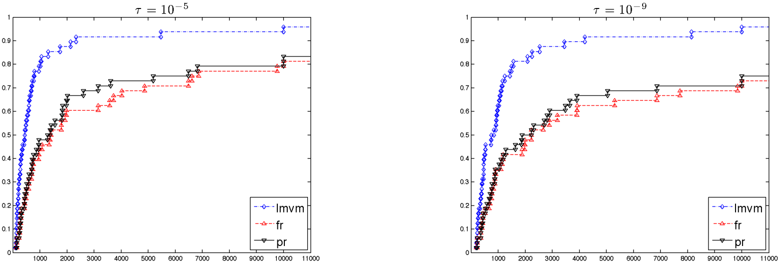
$$\int_{\mathcal{D}} \left\{ -|v(x)|^2 + \frac{1}{2}|v(x)|^4 + \left\| \left[\nabla - iA(x)\right] v(x) \right\|^2 + \kappa^2 \left\| (\nabla \times A)(x) \right\|^2 \right\} dx$$

$v : \mathbb{R}^2 \to \mathbb{C}$ (order parameter)
$A : \mathbb{R}^2 \to \mathbb{R}^2$ (vector potential)



- ◇ Unconstrained problem
- ◇ Non-convex function
- ◇ Unique minimizer, but there is a saddle point
- ◇ Hessian is singular.

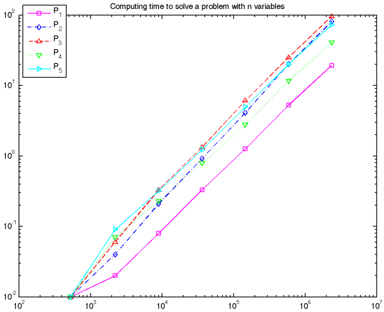
Performance of Nonlinear Conjugate Gradient Methods

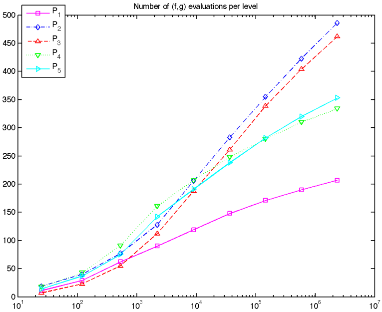


- Data profiles with $f(x) \leq f(x^*) + \tau(f(x_0) - f(x^*))$
- Units are the number of gradients for convergence
- Benchmark problems with n variables, $2,500 \leq n \leq 40,000$

Argonne NATIONAL LABORATORY

Scalability: Computing Time and $(f, \nabla f)$ Evaluations



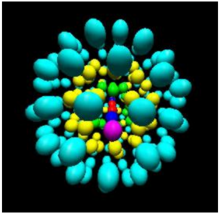


- ◇ Mesh sequencing
- ◇ Convergence test on each level based on f
- ◇ Finest mesh with $1535^2 \approx 2.3\ 10^6$ grid points

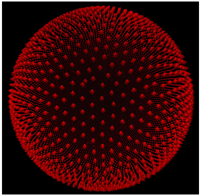
Global Optimization

Determine the optimal configuration $x_1, \ldots, x_m$ in $\mathbb{R}^3$ for

$$V(x) = \sum_{i \neq j} \frac{1}{\|x_i - x_j\|} + \sum_{i=1}^{m} \|x_i\|^2.$$



$m = 200$



$m = 120,000$

Question: Are these optimal configurations?

Concluding Remarks

- Data profiles estimate the percentage of problems that can be solved on a fixed computational budget.
- Model-based derivative free solvers currently offer the best chance of solving noisy, nonlinear optimization problems.
- Seek a professional optimizer for advice on optimization.

Argonne NATIONAL LABORATORY