New Developments with POUNDERS

Stefan Wild

Argonne National Laboratory
Mathematics and Computer Science Division
Joint work with Jorge Moré and Jason Sarich

June 21, 2011

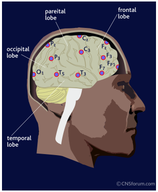
Simulation-Based Optimization (SBO)

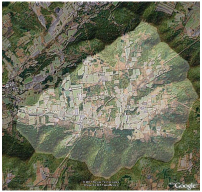
$$\min_{x\in\mathbb{R}^n} \{f(x) = F[S(x)] : c_E(S(x)) = 0, c_I(S(x)) \leq 0\}$$

- x decision variables ("parameters")
- $f : \mathbb{R}^n \to \mathbb{R}$ objective function
- c (vector of) constraints on x
- $S : \mathbb{R}^n \to \mathbb{R}^p$ numerical simulation,
  - typically deterministic and modeling smooth process
- Single evaluation of S expensive

$\Rightarrow$ Evaluation is a bottleneck for optimization

Functions of complex simulations arise far beyond UNEDF





SBO Challenges

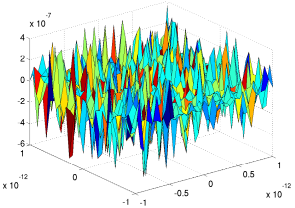
→ Derivatives of S unavailable, prohibitively expensive to approximate directly

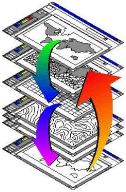
→ Expense increases dramatically as number of parameters increases

→ Computational noise can complicate everything

→ Computational budget limits # S evals

→ Need to exploit parallelism of simulators





→ Obstacles for Automatic Differentiation (coupled legacy/proprietary codes, expense/memory)

Goal: Obtain reduction from $f(x_0)$ in few evaluations

POUNDER

Practical Optimization Using No DERivatives

$\min_x \{f(x) : l \leq x \leq u\}$

- Assumes $\nabla_x f$ unavailable
- Model-based algorithm
- Alternatives:
  - Direct search (Nelder-Mead, Pattern Search)
  - Finite-difference based methods
  - Heuristics (GAs, PSO, SA)
  - Parameter sweeps/hand tuning

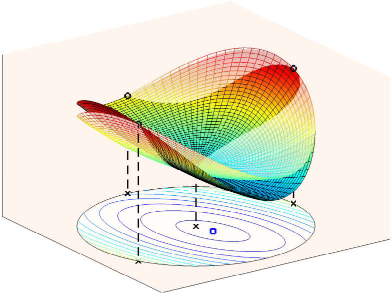
Model-Based Trust Region Algorithms

Substitute $\min\{q_k(x) : x \in \mathcal{B}_k\}$ for $\min f(x)$

$$q_k(x) = f(x_k) + g_k^\top (x - x_k) + \frac{1}{2}(x - x_k)^\top H_k (x - x_k)$$

f expensive, no ∇f?

q_k cheap, analytic derivatives



Update trust region $\mathcal{B}_k$

Based on quality of resulting evaluation

$$\rho_k = \frac{f(x_k) - f(x_+)}{q_k(x_k) - q_k(x_+)}$$

Determine model from known function values

- $q_k(y_i) = f(y_i), \quad \forall y_i \in \mathcal{Y}_k =$ interp. set
- Only need local approximation

$$g_k \approx \nabla f(x_k),\ H_k \approx \nabla^2 f(x_k)$$

- Coarse models $\leftrightarrow$ smooth noise

Adding an S: POUNDERS

Practical Optimization Using No DERivatives (for sums of Squares)

$$\min_x \left\{ f(x) = \sum_i R_i(x)^2 : l \leq x \leq u \right\}$$

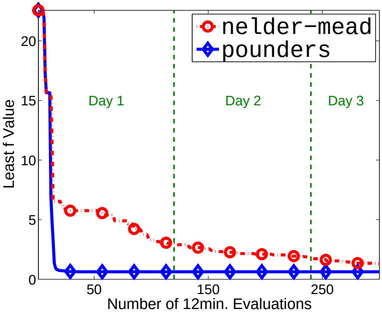
- "Nonlinear least squares"
- Assumes $\nabla_x R_i$ unavailable
- Alternatives:
  - Levenberg-Marquardt with finite-differences
  - Methods for $\min f$
- S also for *Structure*

POUNDERS in UNEDF Calibration of EDFs

With M. Kortelainen, T. Lesinski, J. McDonnell, W. Nazarewicz, N. Schunck, M. Stoitsov (ORNL/UTK)

$$\min_x \left\{ f(x) = \sum_{i=1}^{p} \left(\frac{s_i(x) - d_i}{w_i} \right)^2 \right\}$$

- $s_i(x)$ Simulated nucleus property
- d_i Nucleus i experimental data
- w_i Weight for data type i
- n_{nuc} Parallel simulations



Improved functionals using HFBTHO:

UNEDF0 PRC 2010

UNEDF1 2011

Additional Structure in Simulation-Based Optimization

Removing the UNDER in POUNDERS

Derivatives can be partially available:

- Nonlinear least squares

$$f(x) = \frac{1}{2}\sum_i (s_i(x) - d_i)^2 \qquad \mathbf{\nabla_x s_i(x), i \in I}$$

- Not all variables enter simulation

$$f(x) = g(x_I, x_J) + h(S(x_J)) \qquad \mathbf{\frac{\partial f(x)}{\partial x_i}, i \in I}$$

- Only some constraints depend on S

$$\min\{f(x) : c_1(x) = 0, c_S(x) = 0\} \qquad \mathbf{\nabla_x c_1}$$

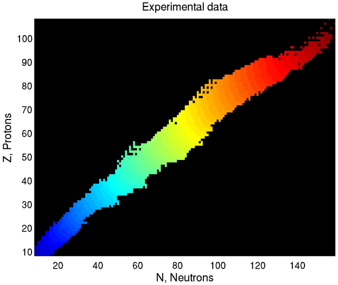
Model-based methods can be a great way to exploit this structure

Multi-level Optimization Structure

With M. Bertolli, T. Papenbrock (ORNL/UTK)

$\min_x \sum_{i=1}^{p} (s_i(x) - d_i)^2$

Fitting $p = 2049$ binding energies



$s_i(x)$ solution to lower level problem

$$\begin{aligned} s_i(x) &= g_i(x) + \min_y \{h_i(x_J; y) : y \in \mathcal{D}_i\} \\ &= g_i(x) + h_i(x_J; y_{i,*}[x_J]) \end{aligned}$$

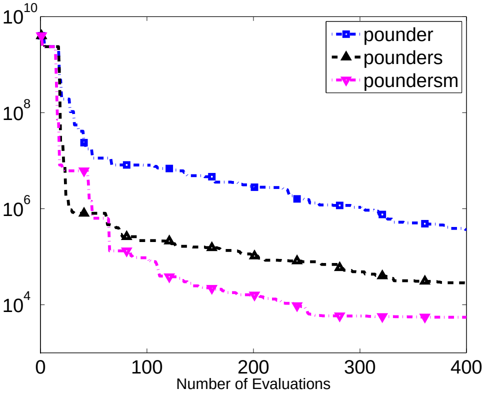
For $x = (x_I, x_J)$

- $\nabla_{x_I} s_i(x_I, x_J)$ available
- $s_i(x)$ continuous and smooth in x_I
- $g_i(x)$ cheap to compute!
- No noise/errors introduced in $g_i(x)$

Numerical Results With Some Derivatives

Bertolli & Papenbrock functional

$$\sum_{i=1}^{p} \left(g_i(x) + h_i(x_J; y_{i,*}[x_J]) - d_i \right)^2$$

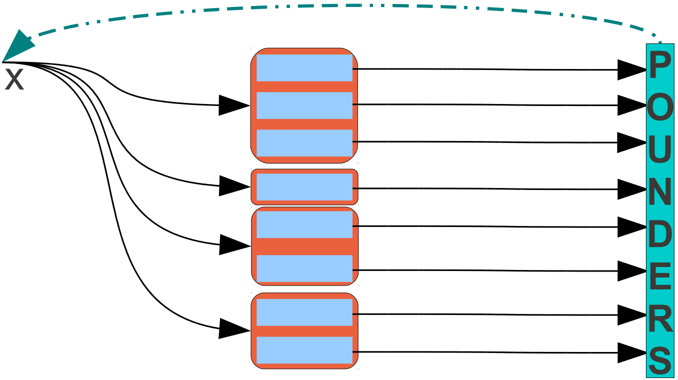


- s exploits least squares
- m uses ∇_{x_I} derivatives

- ◇ $n = 16$ parameters, $|I| = 3$ with derivatives
- ◇ 2049 nuclei in each evaluation, $\mathcal{O}$(secs)

Minimal Output Needed

to benefit from improvements in POUNDERS



- Parallelism of simulators leads to multiplicative scaling
- More on code formatting/scaling from Jason Sarich this afternoon

Summary

Best to go to deepest blackbox level possible:

- Model residuals $\{r_i(x)\}_i$, not $\|r_i(x)\|$
- Model constraints $\{c_i(x)\}_i$, not a penalty $P(c(x))$

$\rightarrow$ POUNDERS allows for incorporating increasing levels of knowledge

Looking for problems where

- Derivatives of only some residuals available
- Simulations require petaflops, noisy
- Lower level problems can be solved inaccurately

Significant Momentum for Optimization Collaborations

Appeared

2009 *Solution of the Skyrme-Hartree-Fock-Bogolyubov equations in the Cartesian deformed harmonic-oscillator basis. (VI) HFODD (v2.38j): a new version of the program*, J. Dobaczewski, W. Satula, B.G. Carlsson, J. Engel, P. Olbratowski, P. Powalowski, M. Sadziak, **J. Sarich**, N. Schunck, A. Staszczak, M. Stoitsov, M. Zalewski, H. Zdunczuk, CPC.

2009 *Towards The Universal Nuclear Energy Density Functional,* M. Stoitsov, **J. Moré**, W. Nazarewicz, J.C. Pei, **J. Sarich**, N. Schunck, A. Staszczak, **S.M. Wild**, JPCS.

2010 *One-quasiparticle States in the Nuclear Energy Density Functional Theory,* N. Schunck, J. Dobaczewski, J. McDonnell, **J. Moré**, W. Nazarewicz, **J. Sarich**, M.V. Stoitsov, PRC.

2010 *Nuclear Energy Density Optimization,* M. Kortelainen, T. Lesinski, **J. Moré**, W. Nazarewicz, **J. Sarich**, N. Schunck, M.V. Stoitsov, **S.M. Wild**, PRC.

Submitted

- *Advancing Nuclear Physics Through TOPS Solvers and Tools*, E Ng, **J. Sarich**, **S.M. Wild**, T. Munson, H. Aktulga, C. Yang, P. Maris, J.P. Vary, M. Kortelainen, W. Nazarewicz, T. Papenbrock, N. Schunck, M.V. Stoitsov, M.G. Bertolli.
- *Computing Heavy Elements,* N. Schunck, A. Baran, M. Kortelainen, J. McDonnell, **J. Moré**, W. Nazarewicz, J. Pei, **J. Sarich**, J. Sheikh, A. Staszczak, M. Stoitsov, **S.M. Wild**.
- *UNEDF: Advanced Scientific Computing Transforms the Low-Energy Nuclear Many-Body Problem* M. Stoitsov, H. Nam, W. Nazarewicz, A. Bulgac, J.C. Pei, I. Thompson, N. Schunck, J.P. Vary, **S.M. Wild**.

Nearing completion

- *"UNEDF1"*, M. Kortelainen, J. McDonnell, W. Nazarewicz, **J. Sarich**, N. Schunck, M.V. Stoitsov, **S.M. Wild**.
- *Occupation-number-based energy functional for nuclear masses*, M. Bertolli, T. Papenbrock, **S.M. Wild**.

Thank You!

Oak Ridge National Laboratory: M. Kortelainen, T. Lesinski, W. Nazarewicz, J. Pei, N. Schunck, A. Staszczak, and M. Stoitsov

Oak Ridge National Laboratory: M. Bertolli and T. Papenbrock

Iowa State University: P. Maris, A. Negoita, A. Shirokov, and J. Vary

Universitas Varsoviensis: G. Carlsson and J. Dobaczewski

Always collecting SBO problems:

→ www.mcs.anl.gov/~wild/sbo

Advance SBO – Attract Optimizers – Find Better Solutions – Fame/Anonymity