Toward Advances in 3 Areas

Dept. 9211 Review, June 2005

David M. Gay (dmg)

Sandia National Laboratories

`dmgay@sandia.gov`

`http://endo.sandia.gov/~dmgay`

2005-3648P

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000

Sandia National Laboratories

Outline

1. Background
2. Automatic Differentiation
3. Global Optimization
4. Data Pipelining
5. Service

Background

- Interests include
  - Scientific computing in general
  - Smooth optimization in particular
- Good experience using AD with AMPL
  - Gradient, Hessian computaions
    on AMPL expression graphs
- Early experience bounding nonlin. eqn sol’ns
- Now at Sandia ➔ chance to work on large-scale computations.
  - This year, focused on three disparate topics.

Automatic Differentiation CSRF

- Work with Eric Phipps (9233); dmg = PI
  - With helpful comments from Roscoe Bartlett, Andy Salinger, etc.
- Goal: promote better use of computing resources via better algorithms for derivatives.
- Work this past year includes
  - Templating Charon (Eric Phipps)
  - Improving, templating RAD (dmg)

Sandia National Laboratories

Often need derivatives

Discretizations often yield nonlinear equations

– To solve f(x) = 0, need Jacobian matrix f'(x)

Optimization: to minimize f(x), want gradient of f

Automatic Differentiation is

– more accurate

– often faster

than finite differences.

In particular, reverse mode AD provides an efficient way to compute the gradient of f.

Action on a Mesh

- In large-scale computation, values of interest are often computed on elements of a mesh.
- Often a small number of mesh-element functions.
- Absent time stepping,

  Visiting each element once and

  computing its contribution to f, rf, r^2f and

  Manually assembling the overall f

may be much more efficient than straightforward than backward AD.

RAD: Specialized overloading for *rf*

- "Active" variables of type ADvar
  - Simply assigned values initially
  - Can be overwritten
  - Each operation stores values, partials in memory
- Invoke

  ```
  ADcontext::Gradcomp();
  ```

  when gradient desired.
  - Causes reverse sweep and reclamation of memory
  - For ADvar v,
    - `v.val()` = current value
    - `v.adj()` = adjoint w.r.t. last computed ADvar value
- Block memory allocation for efficiency.

Example: mesh optimization

$$\tau = det(AW^{-1})$$

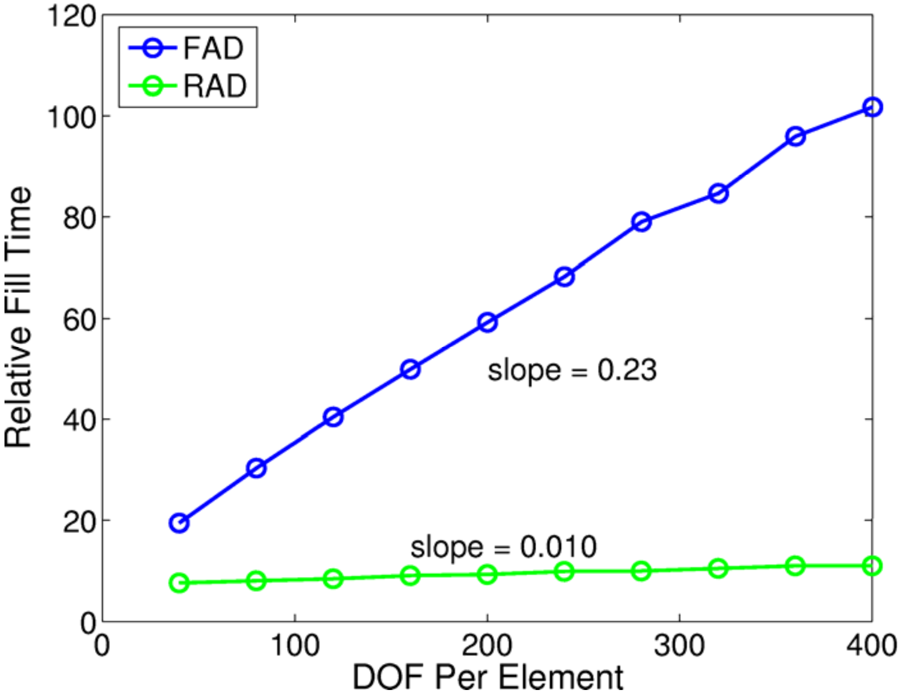
$$h = 0.5(\tau + \sqrt{\tau^2 + 4\delta^2})$$

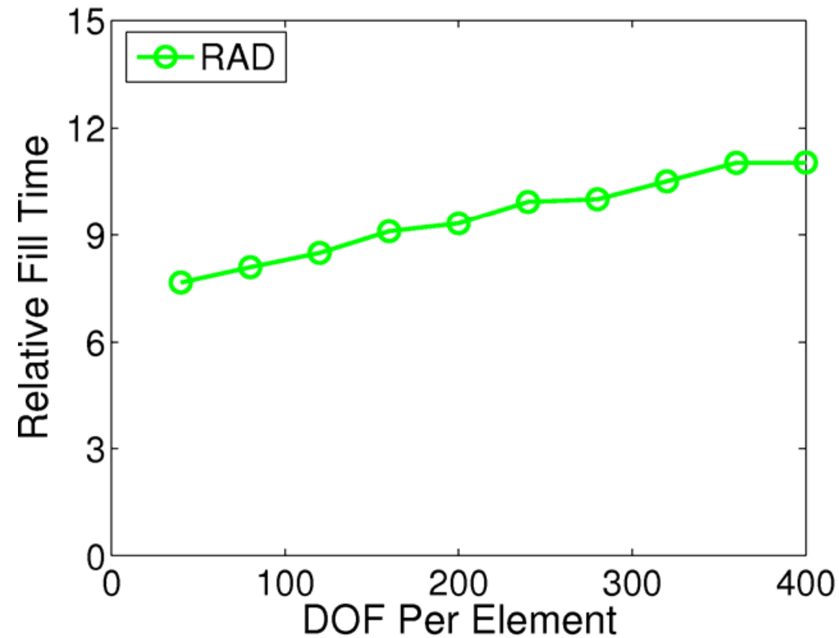
$$\mu_1 = \frac{\|AW^{-1} - I\|_F^2}{h^{2/3}}$$

Relative Times (*f* + r*f*) for *f* = μ_1

	Desktop	Laptop
Compiled *f*	1.07	1.12
RAD	9.14	7.06
ADOL-C new tape	55.0	37.7
ADOL-C old tape	15.4	14.1
C from nlc	1.00	1.00

FAD and RAD in Charon for $\nabla f(x)$





Reverse AD gives gradients or sensitivities for many variables in time proportional to a single function evaluation.

RAD improvements

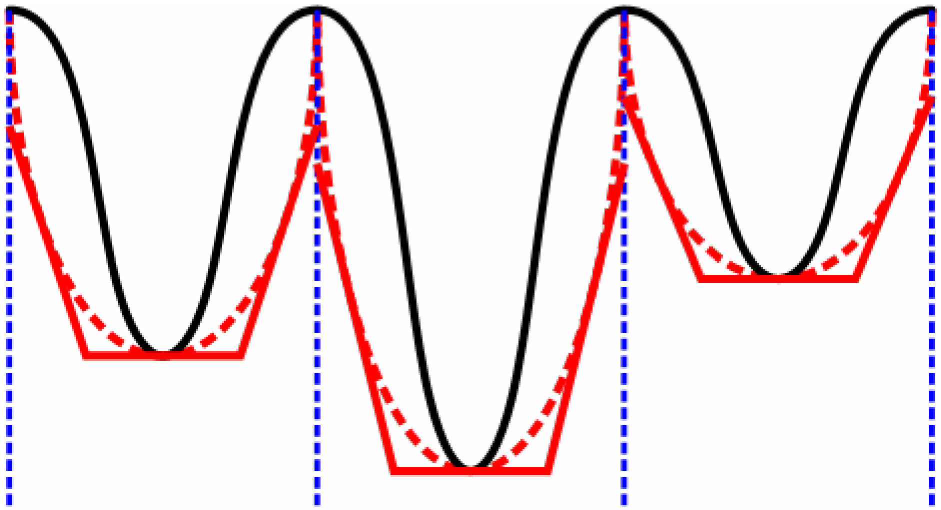
- Templating (for use with various types)
  – May allow combining with TFAD for Hessian-vector products

Motivated by use in Charon:

- Debugging aid via signaling NaNs
- Recording of independent variables
  – and looping over them
- Means to designate "constants" on the fly

Rigorous Global Optimization

PI: Bill Hart
Application example: NW safing



FY05 Rigorous Global Optimization

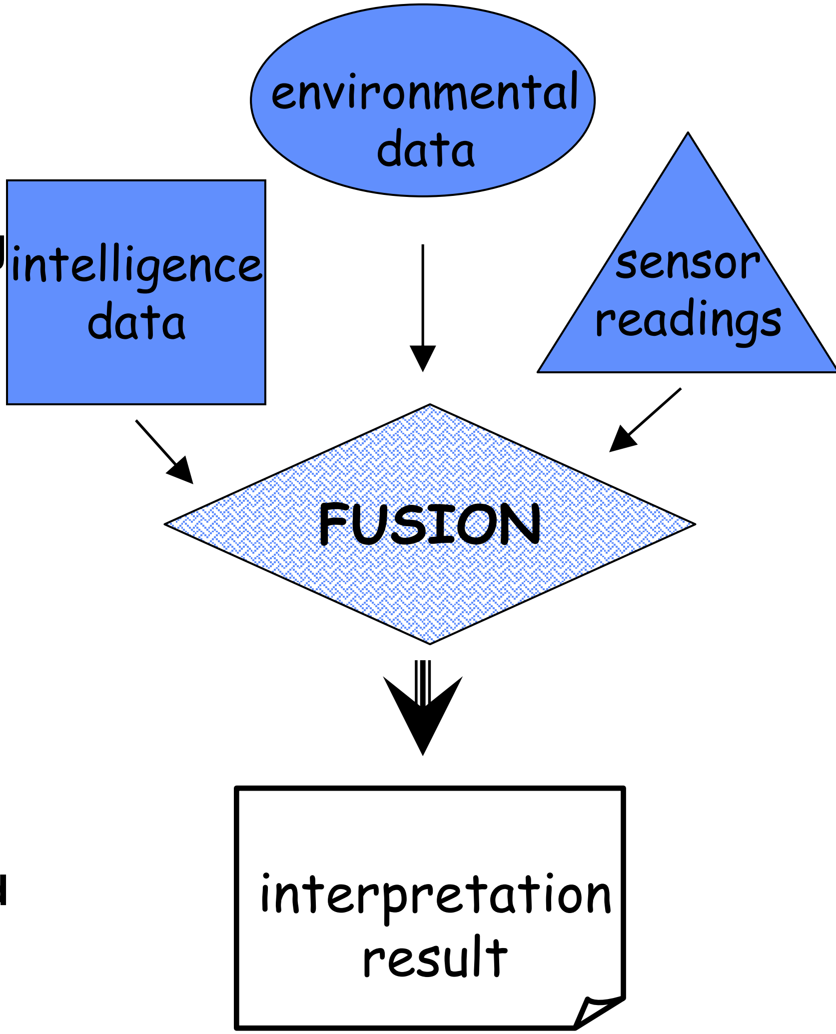
- University "Coconut" global optimization code
  - it's working enough to extract pieces we can use
    - e.g., interval slope computations
  - added driver for NPSOL
  - found OPT++ design limitation and fix

- Bill Hart, PI
- Dave Gay
- Hosting two summer students
  - Lianjun Jiang (U of CO)
    - interval slopes, α-BB for bounding
  - Pradeep Polisetty (U of S. Carolina)
    - convex underestimators from expression graph

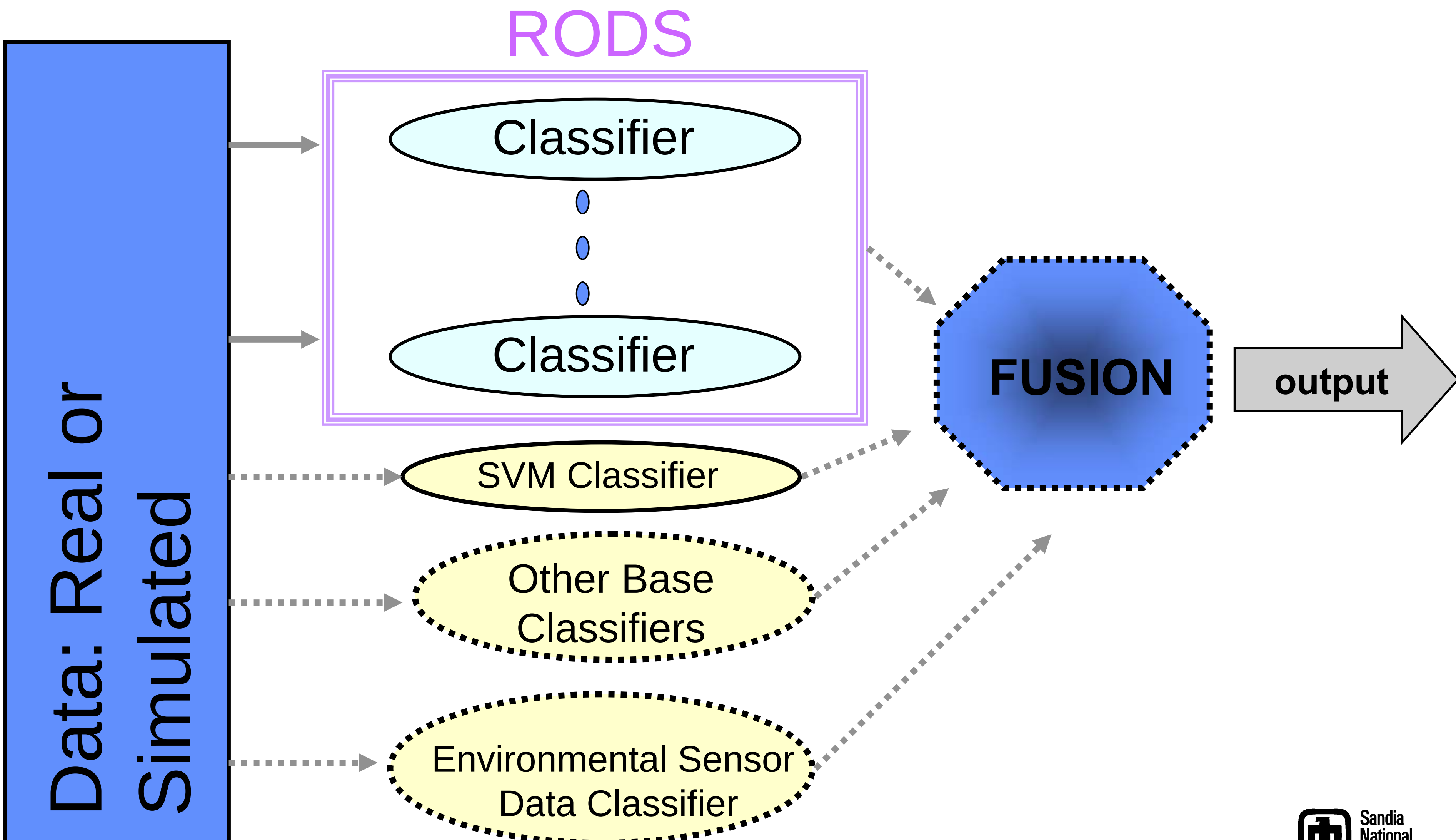
Data Pipelining for Heterogeneous Data Fusion

Project Overview

- **Goal: detection and assessment for chem, bio, and nuclear attacks**
- **Assume a family of "classifiers" monitoring disparate data and providing conclusions**
  - **e.g. hospital records ⇨ high admission rates**
  - **e.g. chemical sensors ⇨ high radiation levels**
  - **e.g. intelligence data ⇨ increased terrorist activity**
- Technical Challenge: **Algorithms for real-time interpretive data mining**
  - **Combine the base classifiers for meta-conclusion**
    - **e.g. dirty bomb attack may have occurred downtown**
    - **goal is subtle scenarios**

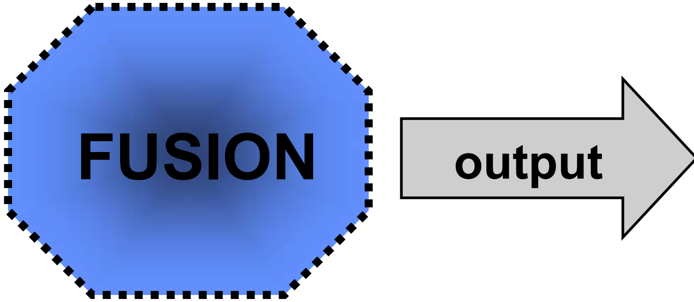


Data pipeline



Data Fusion

- Move from human to automated fusion
- Extend the use of ensemble classifiers to disparate data sources
- Use optimization to provide mathematical rigor & obtain a provable method of solution
- Provide a useful method within the framework



Sandia National Laboratories

Baseline approach = Voting

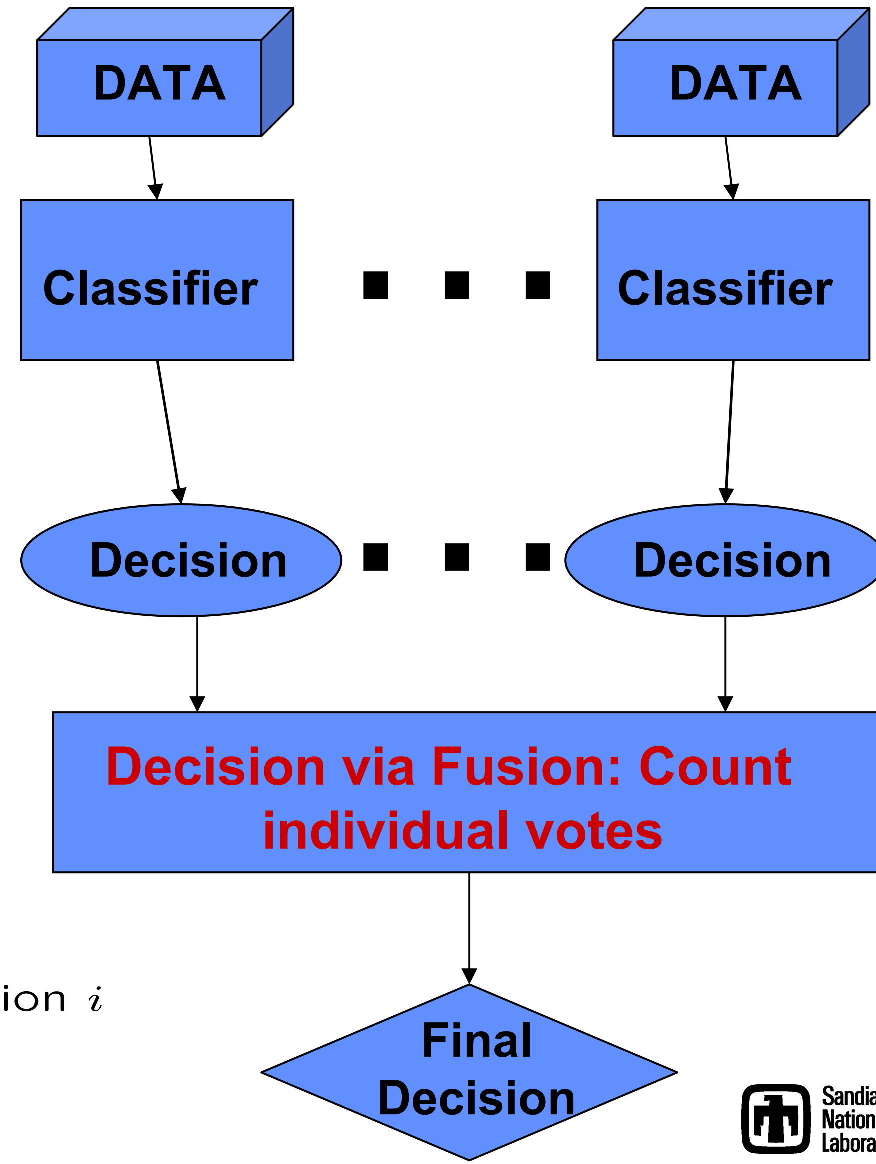
- **Unweighted majority rule**
- **Weighted voting**
  - **Optimize to determine weights, based on training set accuracy**
  - **Simple 8-line AMPL program!**

$$\min \sum_{i=1}^{m} \left(\sum_{j=1}^{n} E_{ij} x_j \right)^2$$

$$\text{s.t.} \quad \sum_{j=1}^{n} x_j = 1, x \geq 0$$

$n =$ number of classifiers
$m =$ number of observations
$E = m \times n$ matrix
$E_{ij} = 1$ if classifer j is wrong in observation i

- **Future: more sophisticated classifiers**



Data Fusion Project Team

2005 LDRD

- Genetha Gray (8962)
  - Data fusion, non-RODS classifiers, PI
- Ken Sale (8321)
  - RODS medical data set classifiers
- Dave Gay (9211)
  - Optimal data fusion for meta-classifier
  - Theory of convergence

- Consultants
  - Heidi Ammerlahn (8112 → 8962)
  - Keith Vanderveen (8112)
  - Pam Williams (8962)

Service

- Math. Programming Society Treasurer
- AMPL support
  R&D and international support at home
  SNL site license
    Julie Lloyd (6221), Vitus Leung
    Bob Carr
    van Bloemen Waanders, water security team
- Refereeing
  - Two AD04 proceedings papers
  - Papers for Math. Programming, SIAM J Optim.

Automatic Differentiation (AD) CSRF Project Funded for FY05-07

- Acquired funding for a new CSRF project on AD with David Gay (PI, 9211)
- Develop AD tools for Sandia DP applications
  - Forward: Extend and "robustify" TFAD
  - Reverse: Investigate efficiency, scalability of operator overloading approaches
  - Taylor: Study Taylor polynomial based algorithms for time integration, uncertainty quantification
- Goals:
  - Make AD a commonplace technology at Sandia
  - Provide derivative technology to foster new algorithm development

Finite element residual equations

$$f(x) = 0$$

$$\frac{\partial f}{\partial x} V$$

$$\left(\frac{\partial f}{\partial x}\right)^T V$$

$$x(t) = \sum_{k=0}^{d} x_k t^k$$

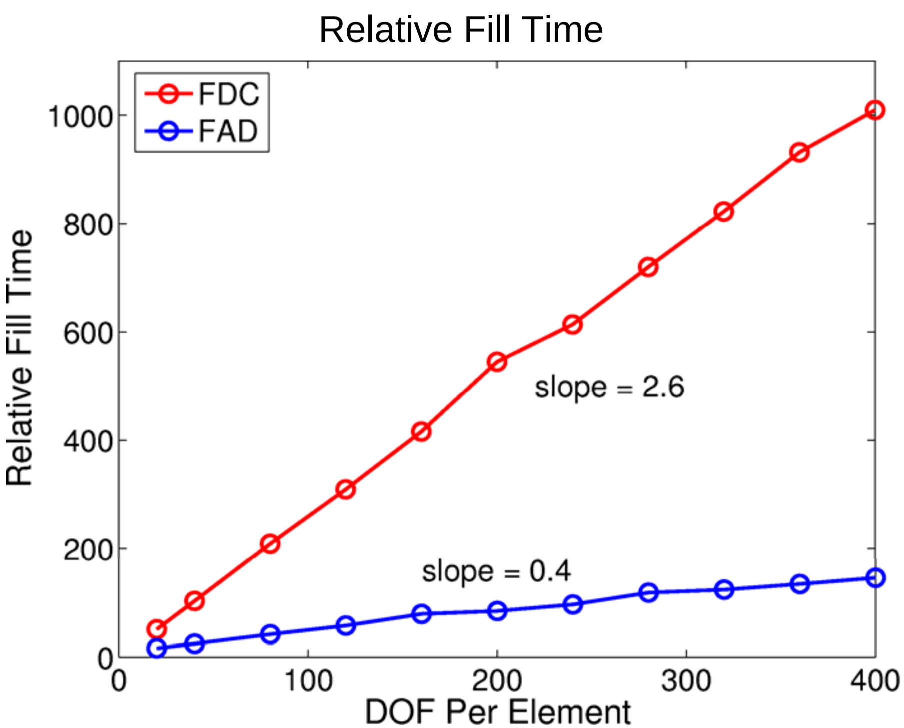
$$f(x(t)) = \sum_{k=0}^{d} f_k t^k + O(t^{d+1})$$

"Analytic derivatives are critical for the success of the Charon/QASPR project." – Rob Hoekstra, Charon Project Lead

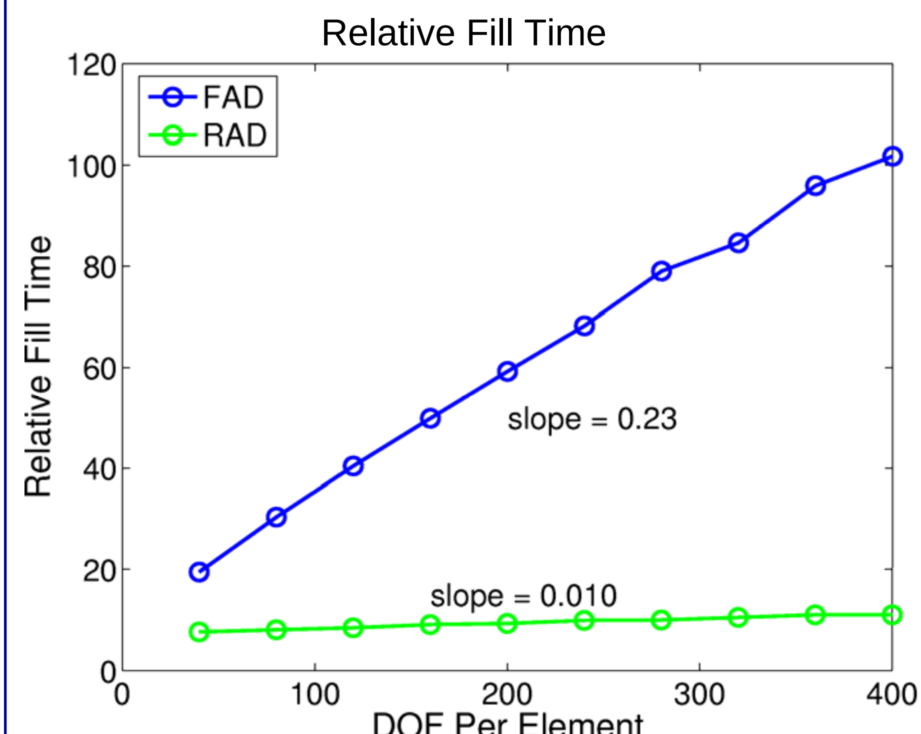
Sandia National Laboratories

Forward and Reverse Mode AD in Charon

Forward Mode Jacobian Using FAD Versus Finite Differencing w/Coloring



Reverse Mode Jacobian-Transpose Product Versus Forward Mode



FAD has become the standard Jacobian strategy in Charon

Enabling technology for large-scale sensitivity analysis & optimization

Sandia National Laboratories