How to Choose an Algorithm

Matthew Knepley

Computer Science and Engineering
University at Buffalo

Finger Lakes HPC Meeting
University of Rochester
Rochester, NY October 14, 2022

Argonne NATIONAL LABORATORY

University at Buffalo

RELACS People

How do we choose an algorithm?

How do we choose an algorithm?

We choose the fastest one…

Timing is tricky. It’s sensitive to

Timing is tricky. It’s sensitive to

machine characteristics

Timing is tricky. It’s sensitive to

machine characteristics

problem details

Proxy measures can simplify design:

Proxy measures can simplify design:

Computation (HPL)

Proxy measures can simplify design:

Computation	(HPL)
Bandwidth	(Roofline)

Proxy measures can simplify design:

Computation	(HPL)
Bandwidth	(Roofline)
Latency	(LogP)

Proxy measures can simplify design:

Computation	(HPL)
Bandwidth	(Roofline)
Latency	(LogP)
Concurrency	

These models can answer…

These models can answer…

Does this implementation scale weakly?

These models can answer…

Does this implementation scale weakly? strongly?

These models can answer…

Is one implementation more efficient than another on this machine?

What about questions like...

What about questions like...

Should I discretize this problem with CG or DG?

What about questions like...

Should I solve using the Picard or Newton Method?

The key notion we are missing is

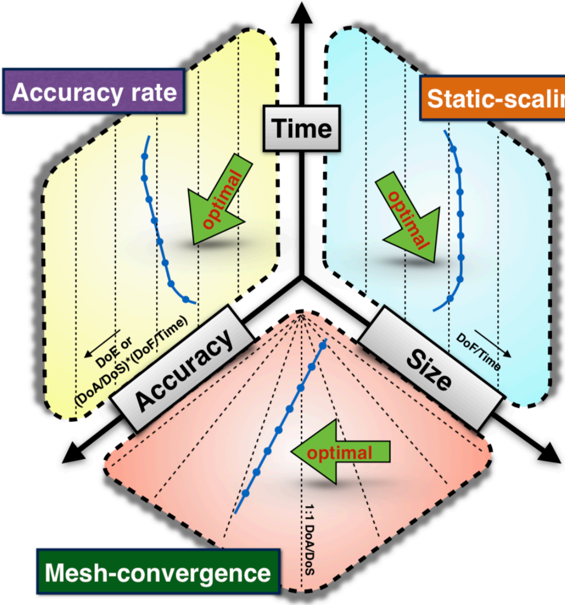
The key notion we are missing is

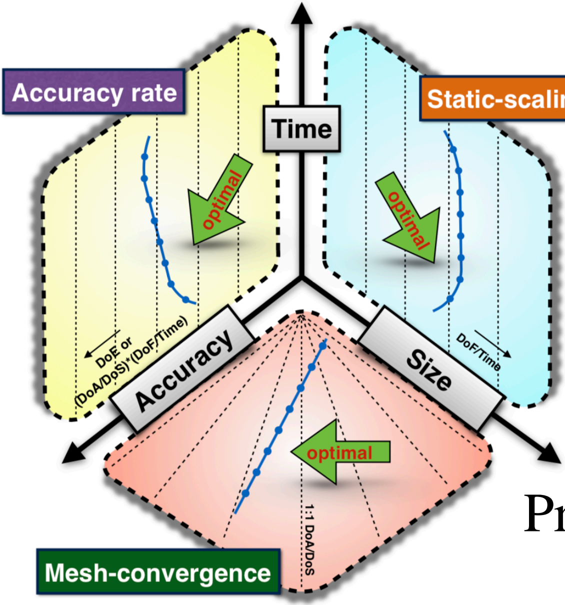
accuracy

The key notion we are missing is

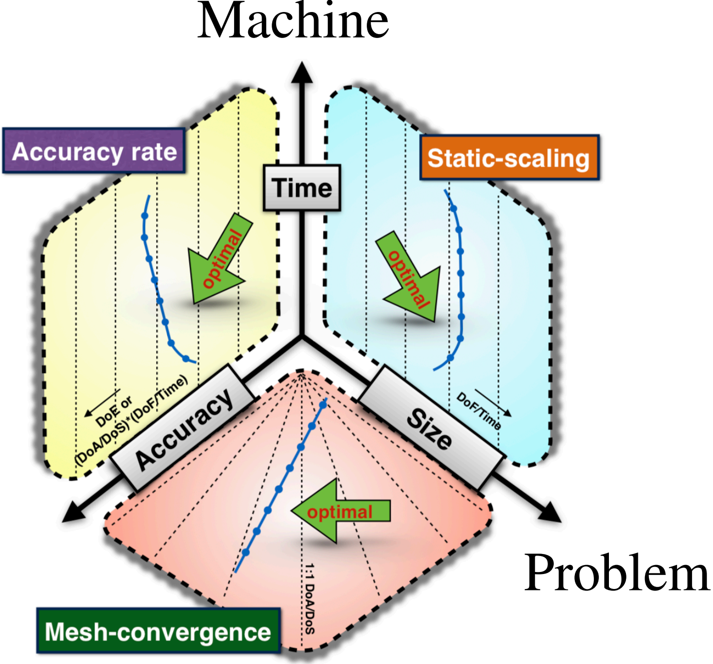
accuracy

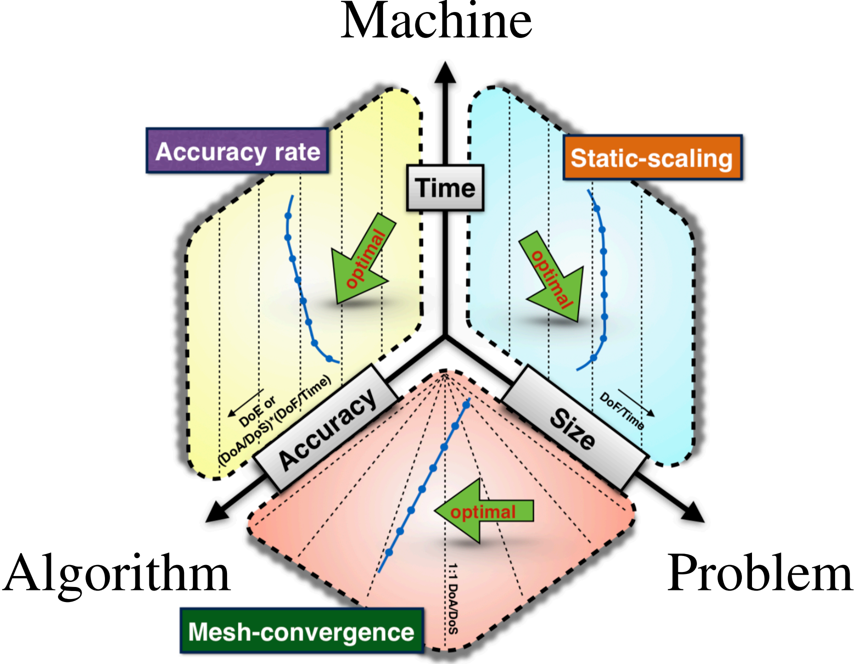
It distinguishes algorithms with different convergence behavior (Chang et al. 2018)


Time
Accuracy rate
Static-scaling
optimal
optimal
optimal
DoE or
(DoA/DoS)*(DoF/Time)
DoF/Time
Accuracy
Size
1:1 DoA/DoS
Mesh-convergence

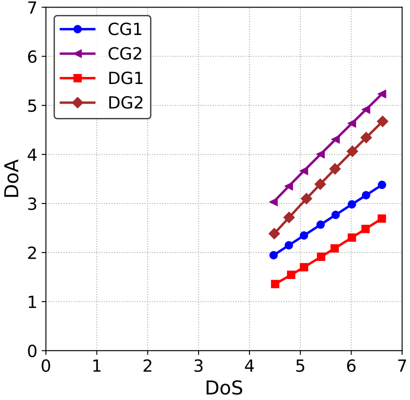

Accuracy rate
Time
Static-scaling
optimal
optimal
DoE or
(DoA/DoS)*(DoF/Time)
Accuracy
Size
DoF/Time
optimal
1:1 DoA/DoS
Mesh-convergence

Problem

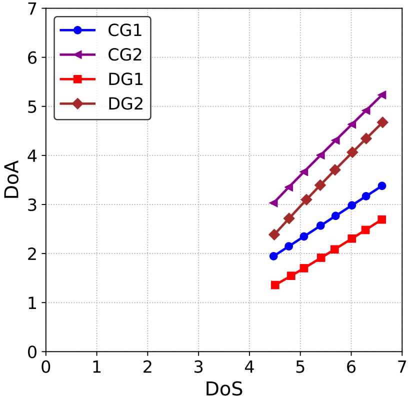

Machine
Accuracy rate
Time
Static-scaling
optimal
optimal
DoE or
(DoA/DoS)*(DoF/Time)
Accuracy
Size
DoF/Time
optimal
1:1 DoA/DoS
Mesh-convergence
Problem


Machine
Accuracy rate
Time
Static-scaling
optimal
optimal
DoE or
(DoA/DoS)*(DoF/Time)
Accuracy
Size
DoF/Time
optimal
1:1 DoA/DoS
Algorithm
Problem
Mesh-convergence

Mesh Convergence Diagram

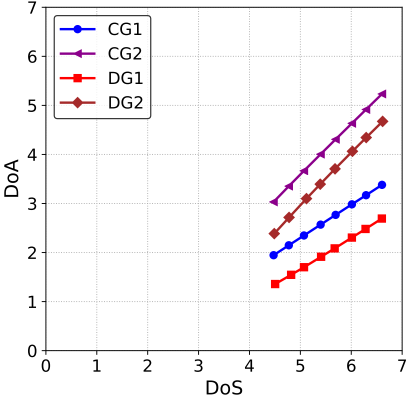


Mesh Convergence Diagram



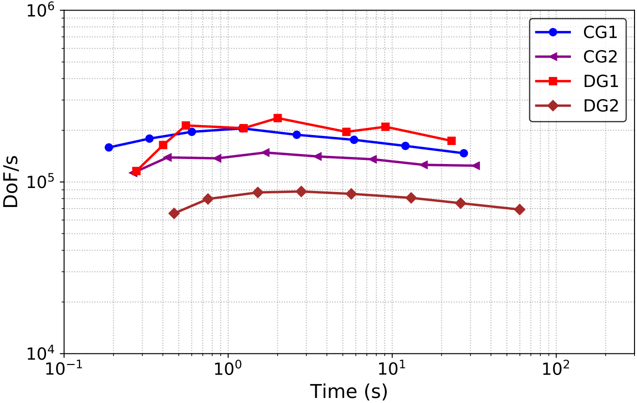
1/error vs. size

Mesh Convergence Diagram

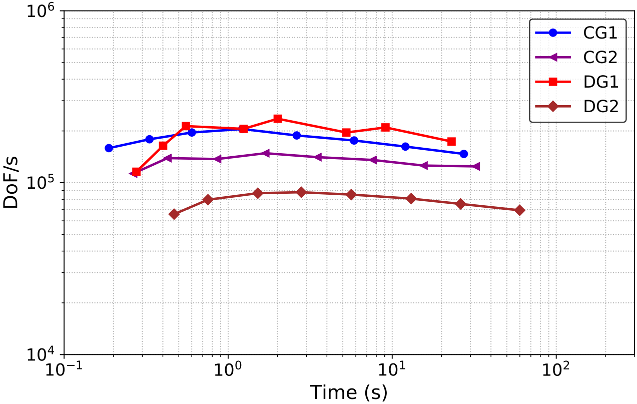


Does my Algorithm solve this Problem?

Static Scaling Diagram

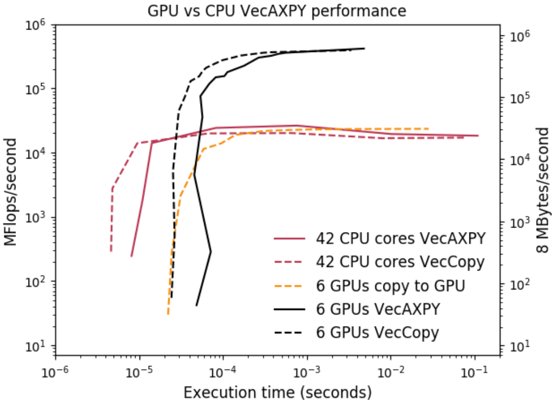


Static Scaling Diagram



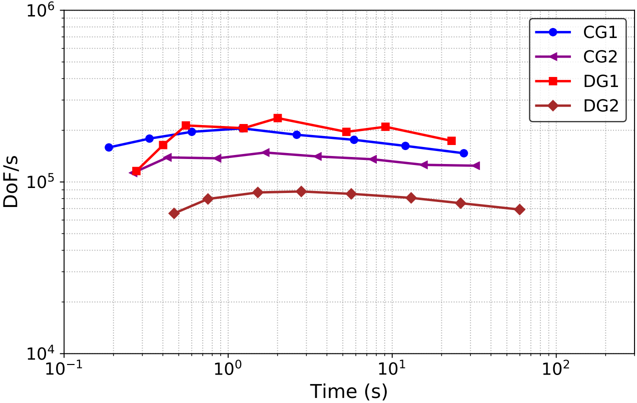
size/time vs. time

Static Scaling Diagram



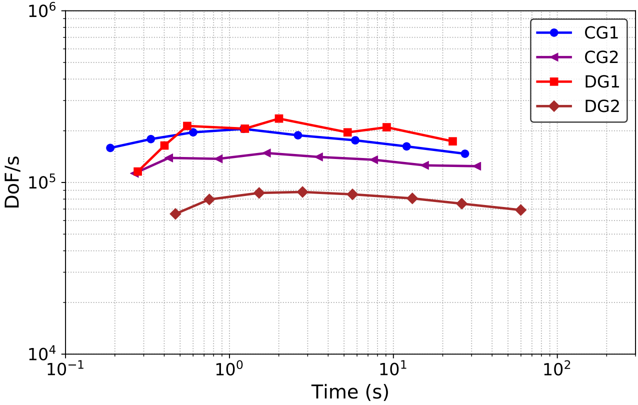
size/time vs. time

Static Scaling Diagram



size/time vs. time

Static Scaling Diagram



Is my Algorithm efficient on this Machine?

How should we measure accuracy?

How should we measure accuracy?

accuracy rate $\frac{e}{T}$

How should we measure accuracy?

accuracy rate $\frac{e}{T}$

Marginal accuracy rate falls off steeply with problem size

Consider an optimal PDE solver:

Consider an optimal PDE solver:

$$T = Wh^{-d} \text{ and } e = Ch^{\alpha}$$

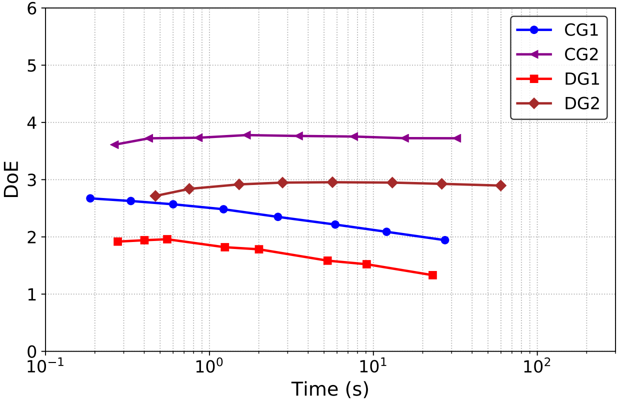
Consider an optimal PDE solver:

$$T = Wh^{-d} \text{ and } e = Ch^{\alpha}$$

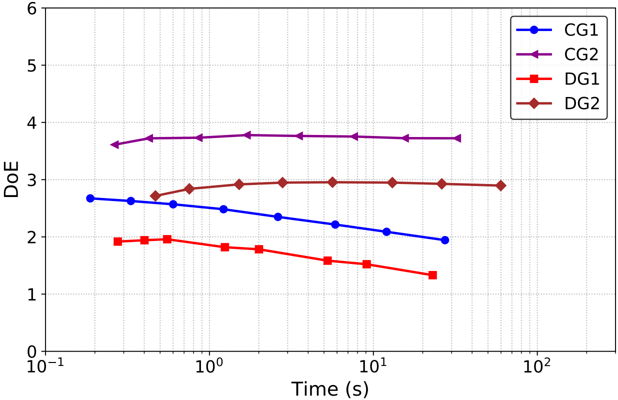
The error-time has a simple form

$$\begin{aligned} & -\log(e \cdot T) \\ = & -\log\left(Ch^{\alpha}Wh^{-d}\right) \\ = & (d-\alpha)\log(h) - \log(CW) \end{aligned}$$

Efficacy Diagram

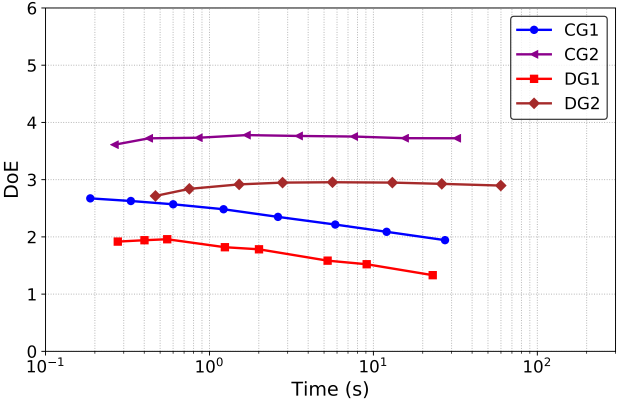


Efficacy Diagram



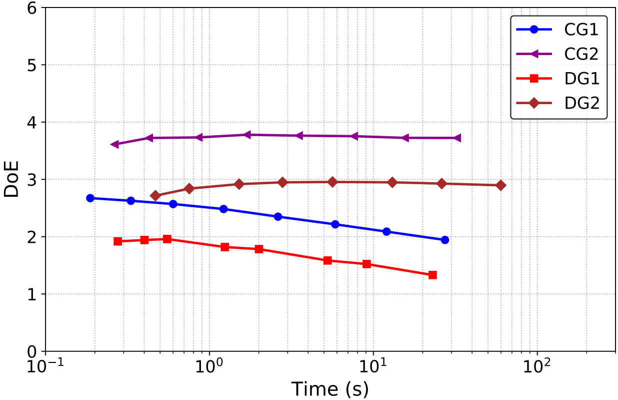
1/error-time vs. time

Efficacy Diagram



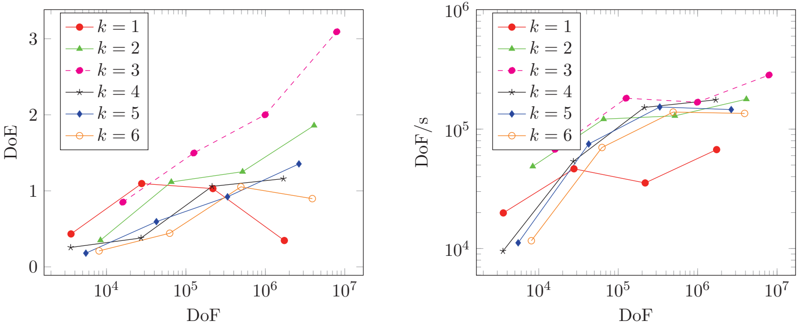
$$\frac{1/\text{error}}{\text{size}} \times \frac{\text{size}/\text{time}}{\text{time}} = \frac{1/(\text{error}\cdot\text{time})}{\text{time}}$$

Efficacy Diagram



Does my Algorithm solve this Problem efficiently on this Machine?

Efficacy vs. Static Scaling



(a) DoE vs DoF

(b) DoF/s vs DoF

Figure 17: Time-accuracy performance analysis for the nearly incompressible problem ($\lambda = 10^6$).

(Fabien 2019)

What else could we analyze?

What else could we analyze?

Communication-Avoiding (CA) algorithms have exciting lower bounds

(Ballard et al. 2011)

What else could we analyze?

CA TSQR is a great success

(Demmel et al. 2012)

What else could we analyze?

CA Krylov not a success

What else could we analyze?

CA Krylov not a success

Accuracy depends on coarse grid communication in preconditioner

Future Questions:

Future Questions:

Is there a variational characterization of optimal algorithms?

Future Questions:

Can we think of error-time as an Algorithmic Action?

References I

Chang, Justin, Maurice S. Fabien, Matthew G. Knepley, and Richard T. Mills (2018). "Comparative study of finite element methods using the Time-Accuracy-Size (TAS) spectrum analysis". In: SIAM Journal on Scientific Computing 40.6, pp. C779–C802. DOI: 10.1137/18M1172260. eprint: 1802.07832.

Fabien, Maurice S. (2019). "A GPU-Accelerated Hybridizable Discontinuous Galerkin Method for Linear Elasticity". In: Communications in Computational Physics 27.2, pp. 513–545. ISSN: 1991-7120. DOI: 10.4208/cicp.OA-2018-0235.

Ballard, Grey, James Demmel, Olga Holtz, and Oded Schwartz (2011). "Minimizing communication in numerical linear algebra". In: SIAM Journal on Matrix Analysis and Applications 32.3, pp. 866–901.

Demmel, James, Laura Grigori, Mark Hoemmen, and Julien Langou (2012). "Communication-optimal parallel and sequential QR and LU factorizations". In: SIAM Journal on Scientific Computing 34.1, A206–A239.