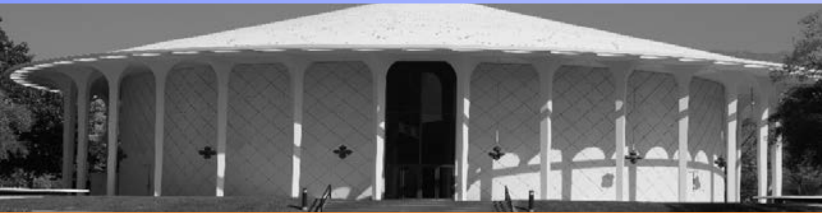
C A L I F O R N I A   I N S T I T U T E   O F   T E C H N O L O G Y



universitätbonn

Data-Driven Computational Mechanics (II)

Michael Ortiz
California Institute of Technology and
Rheinische Friedrich-Wilhelms Universität Bonn

COMPLAS 2017
Barcelona, September 5, 2017

Data Science, Big Data… What's in it for us?

http://olap.com/forget-big-data-lets-talk-about-all-data/

Michael Ortiz
COMPLAS17

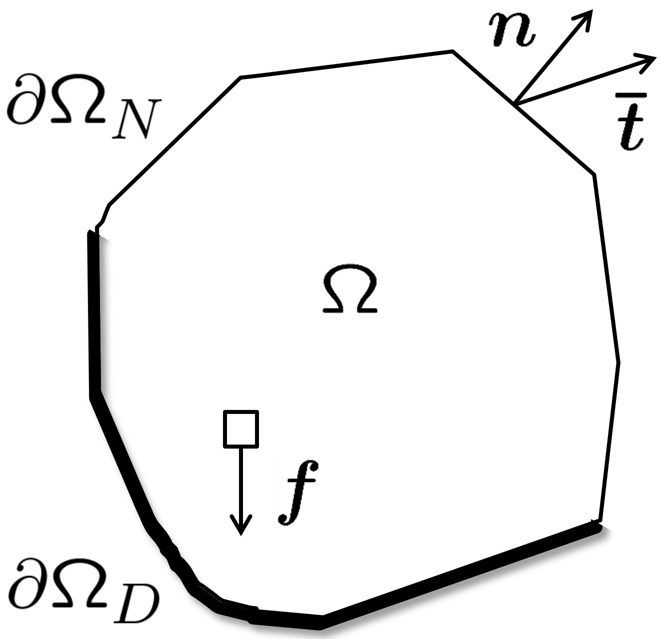
Data Science, Big Data… What's in it for us?

- *Data Science* is the extraction of '*knowledge*' from large volumes of unstructured data[1]
- Data science requires sorting through *big-data* sets and extracting '*insights*' from these data
- Data science uses data management, statistics and machine learning to derive *mathematical models* for subsequent use in decision making
- Data science influences (*non-STEM*!) fields such as marketing, advertising, finance, social sciences, security, policy, medical informatics...
- *But... What's in it for us?*

[1]Dhar, V., *Communications of the ACM*, **56**(12) (2013) p. 64.

Where is Data Science needed in Computational Mechanics?

- Anatomy of a field-theoretical *STEM* problem:



i) Kinematics + Dirichlet:

$$\left.\begin{aligned} &\epsilon(\boldsymbol{u}) = 1/2(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T) \\ &\boldsymbol{u} = \bar{\boldsymbol{u}}, \quad \text{on } \partial\Omega_D \end{aligned}\right\}$$

ii) Equilibrium + Neumann:

$$\left.\begin{aligned} &\operatorname{div} \boldsymbol{\sigma} + \boldsymbol{f} = \boldsymbol{0} \\ &\boldsymbol{\sigma n} = \bar{\boldsymbol{t}}, \quad \text{on } \partial\Omega_N \end{aligned}\right\}$$

iii) Material law: $\boldsymbol{\sigma}(\boldsymbol{x}) = \boldsymbol{\sigma}\Big(\epsilon(\boldsymbol{x})\Big)$

Where is Data Science needed in Computational Mechanics?

- Anatomy of a field-theoretical *STEM* problem:

Universal laws!
(Newton's laws,
Schrodinger's eq.,
Maxwell's eqs.,
Einstein's eqs...)
Exactly known!
Uncertainty-free!
(epistemic)

i) Kinematics + Dirichlet:

$$\left.\begin{array}{l}\epsilon(\boldsymbol{u}) = 1/2(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T) \\ \boldsymbol{u} = \bar{\boldsymbol{u}}, \quad \text{on } \partial\Omega_D\end{array}\right\}$$

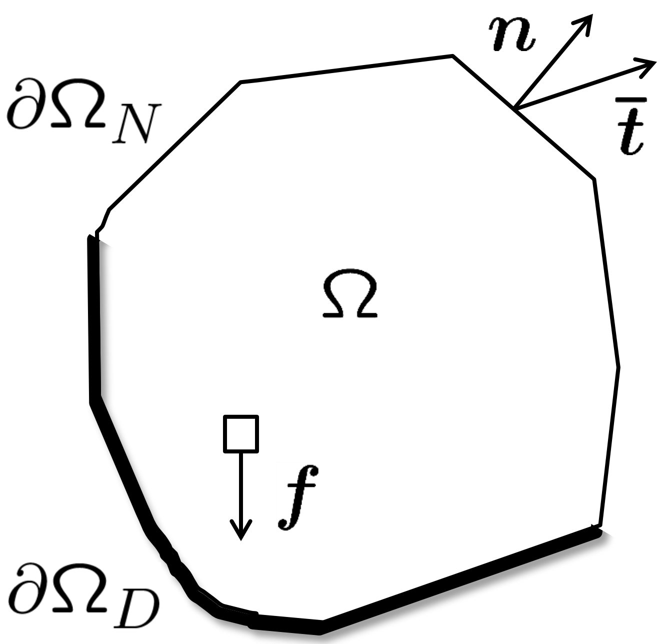
ii) Equilibrium + Neumann:

$$\left.\begin{array}{l}\text{div}\, \boldsymbol{\sigma} + \boldsymbol{f} = \boldsymbol{0} \\ \boldsymbol{\sigma}\boldsymbol{n} = \bar{\boldsymbol{t}}, \quad \text{on } \partial\Omega_N\end{array}\right\}$$

iii) Material law: $\boldsymbol{\sigma}(\boldsymbol{x}) = \boldsymbol{\sigma}\Big(\epsilon(\boldsymbol{x})\Big)$

Where is Data Science needed in Computational Mechanics?

- Anatomy of a field-theoretical *STEM* problem:



i) Kinematics + Dirichlet:

$$\left.\begin{aligned} &\epsilon(\boldsymbol{u}) = 1/2(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T) \\ &\boldsymbol{u} = \bar{\boldsymbol{u}}, \quad \text{on } \partial\Omega_D \end{aligned}\right\}$$

ii) Equilibrium + Neumann:

$$\left.\begin{aligned} &\text{div}\, \boldsymbol{\sigma} + \boldsymbol{f} = \boldsymbol{0} \\ &\boldsymbol{\sigma}\boldsymbol{n} = \bar{\boldsymbol{t}}, \quad \text{on } \partial\Omega_N \end{aligned}\right\}$$

iii) Material law: $\boldsymbol{\sigma}(\boldsymbol{x}) = \boldsymbol{\sigma}\big(\epsilon(\boldsymbol{x})\big)$

Unknown! Epistemic uncertainty!

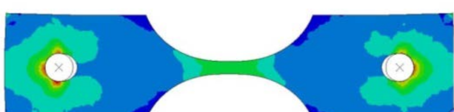
Michael Ortiz
COMPLAS17

Data Science and material modeling

- Need to generate (epistemic) ‘*knowledge*’ about material behavior to close BV problems...
- Traditional *modeling paradigm*: Fit data (a.k.a. regression, machine learning, model reduction, central manifolds...), use calibrated empirical models in BV problems

Data Science and the classical Modeling & Simulation paradigm

$$\sigma = \mathbb{C}\epsilon$$



Material data

Modeling funnel

Material model

funnel Simulation

Manufactured data

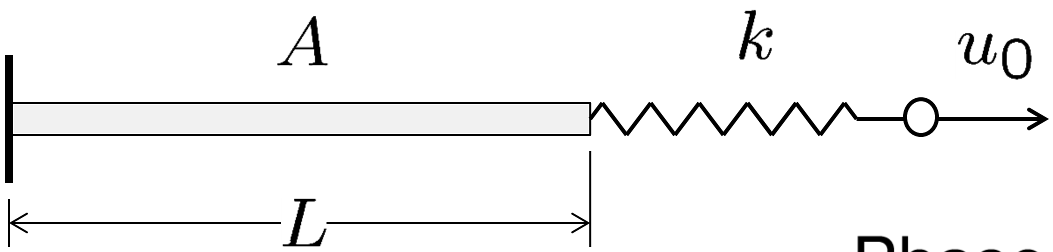
broken pipe!

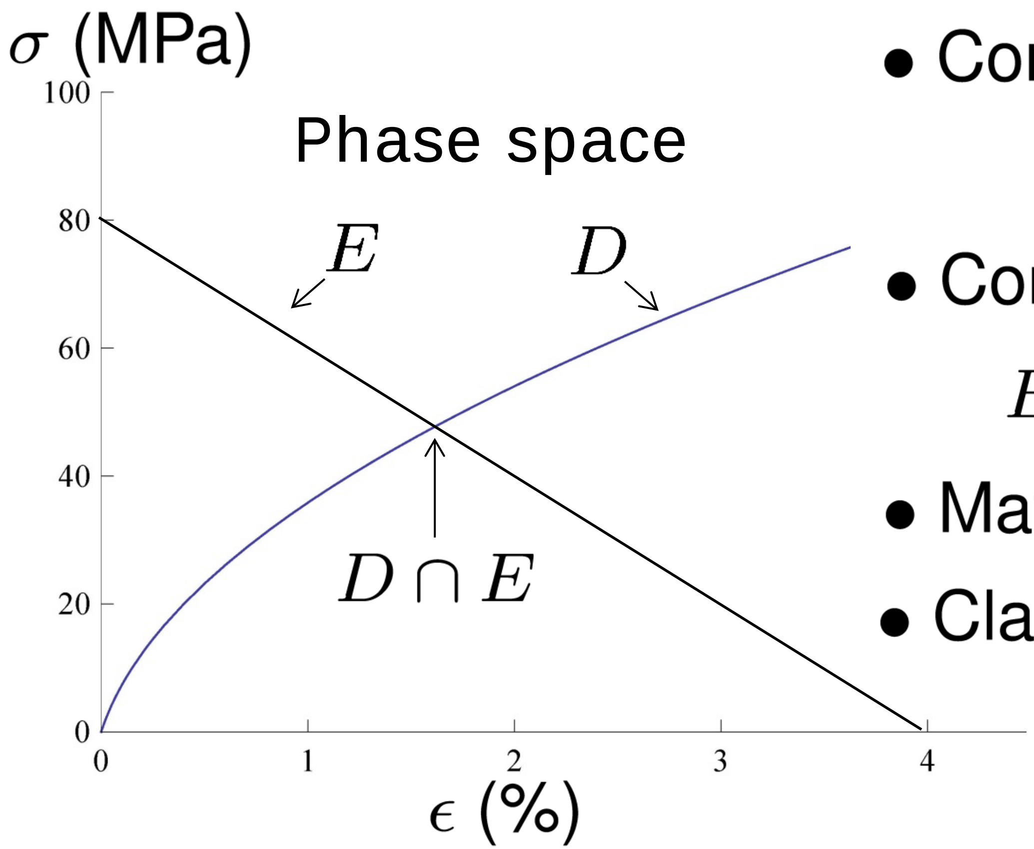
Michael Ortiz
COMPLAS17

Data Science and material modeling

- Need to generate (epistemic) ‘*knowledge*’ about material behavior to close BV problems...
- Traditional *modeling paradigm*: Fit data (a.k.a. regression, machine learning, model reduction, central manifolds...), use calibrated empirical models in BV problems
- *But:* We live in a *data-rich world* (full-field diagnostics, data mining from first principles...)
- *Data-Driven paradigm*: Use material data directly in BV (no fitting by any name, no loss of information, no broken pipe between material data and manufactured data)
- *How?* (math talks, nonsense walks...)

Elementary example: Bar and spring





- Phase space: $\{(\epsilon, \sigma)\} \equiv Z$
- Compatibility + equilibrium:

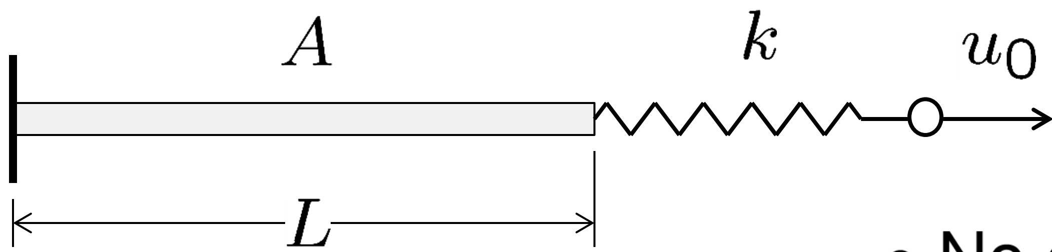
$$\sigma A = k(u_0 - \epsilon L)$$

- Constraint set:

$$E = \{\sigma A = k(u_0 - \epsilon L)\}$$

- Material data set: $D \subset Z$
- Classical solution set: $D \cap E$

Elementary example: Bar and spring

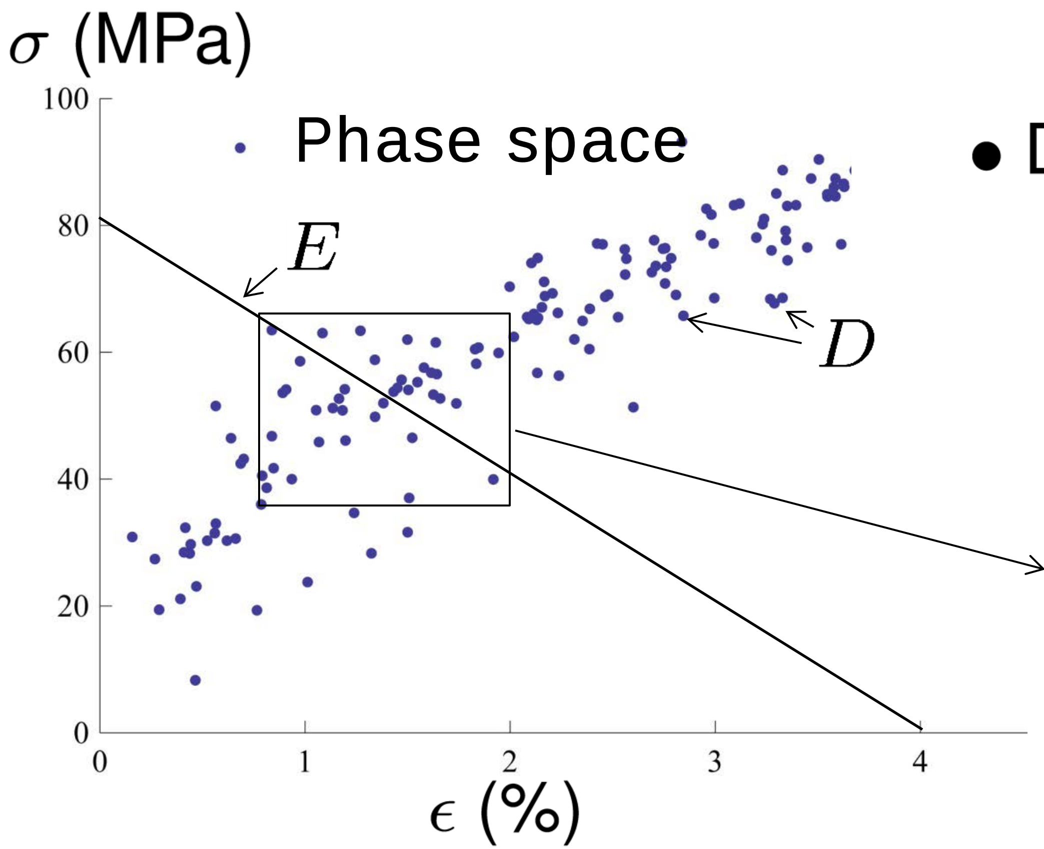


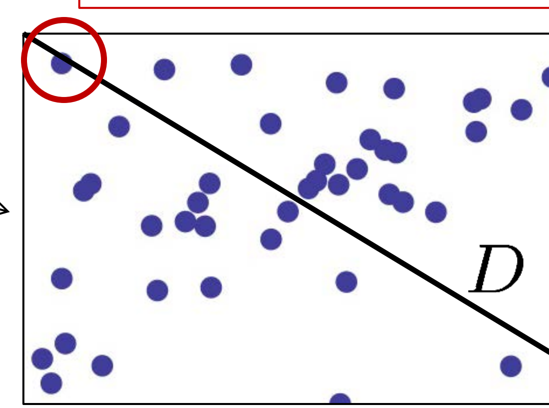
- No classical solutions!

$$D \cap E = \emptyset$$

- Data-driven solution:

$$\min_{z \in E} \operatorname{dist}(z, D)$$





Michael Ortiz
COMPLAS17

The general Data-Driven (DD) problem

- The Data-Driven problem[1]: Given,
  - *D = {material data},*
  - *E = {compatibility + equilibrium},*

$$\text{Find: argmin}\{d(z, D),\ z \in E\}$$

- ***The aim Data-Driven analysis is to find the compatible and equilibrated solution that is closest to the material data set***
- No material modeling, no data fitting, no V&V...
- Raw material data is used (unprocessed) in calculations (the facts, nothing but the facts...)
- ***Are Data-Driven problems well-posed?***

[1]T. Kirchdoerfer and M. Ortiz (2015) arXiv:1510.04232.
[1]T. Kirchdoerfer and M. Ortiz, *CMAME*, **304** (2016) 81–101

Data-Driven elasticity

Definition (Phase space)

$$Z = \{\epsilon \in L^2(\Omega; \mathbb{R}^{n\times n}_{\mathrm{sym}})\} \times \{\sigma \in L^2(\Omega; \mathbb{R}^{n\times n}_{\mathrm{sym}})\}.$$

Definition (Constraint set)

i) Compatibility,

$$\epsilon = 1/2(\nabla u + \nabla u^T),$$

$$u = g, \quad \text{on } \Gamma_D.$$

ii) Equilibrium,

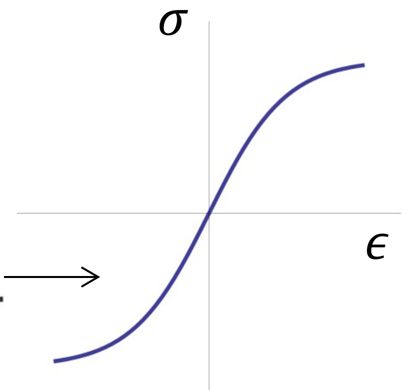
$$\mathrm{div}\sigma + f = 0,$$

$$\sigma\nu = h, \quad \text{on } \Gamma_N.$$

Definition (Material data set)

Hooke's law (linear) $D = \{\sigma = \mathbb{C}\epsilon\}$.

Hooke's law (monotone) $D = \{\sigma = \sigma(\epsilon)\}$. →



S. Conti, S. Müller and M. Ortiz (2017) arXiv:1708:02880

Data-Driven elasticity – Well-posedness

Theorem (Existence and uniqueness)

Let $\Omega \subset \mathbb{R}^n$ *be open, bounded, Lipschitz. Let* $\overline{\Gamma}_D \cap \overline{\Gamma}_N = \partial\Omega$, $\mathcal{H}^{n-1}(\overline{\Gamma}_N \setminus \Gamma_N) = \mathcal{H}^{n-1}(\overline{\Gamma}_D \setminus \Gamma_D) = 0$ *and* $\Gamma_D \neq \emptyset$. *Assume:*

i) $\mathbb{C} \in L(\mathbb{R}^{n\times n}_{\mathrm{sym}})$, $\mathbb{C}^T = \mathbb{C}$, $\mathbb{C} > 0$.

ii) $f \in L^2(\Omega; \mathbb{R}^n)$, $g \in H^{1/2}(\partial\Omega; \mathbb{R}^n)$, $h \in H^{-1/2}(\partial\Omega; \mathbb{R}^n)$.

Then, the Data-Driven problem

$$\min\{d(z, D),\ z \in E\}$$

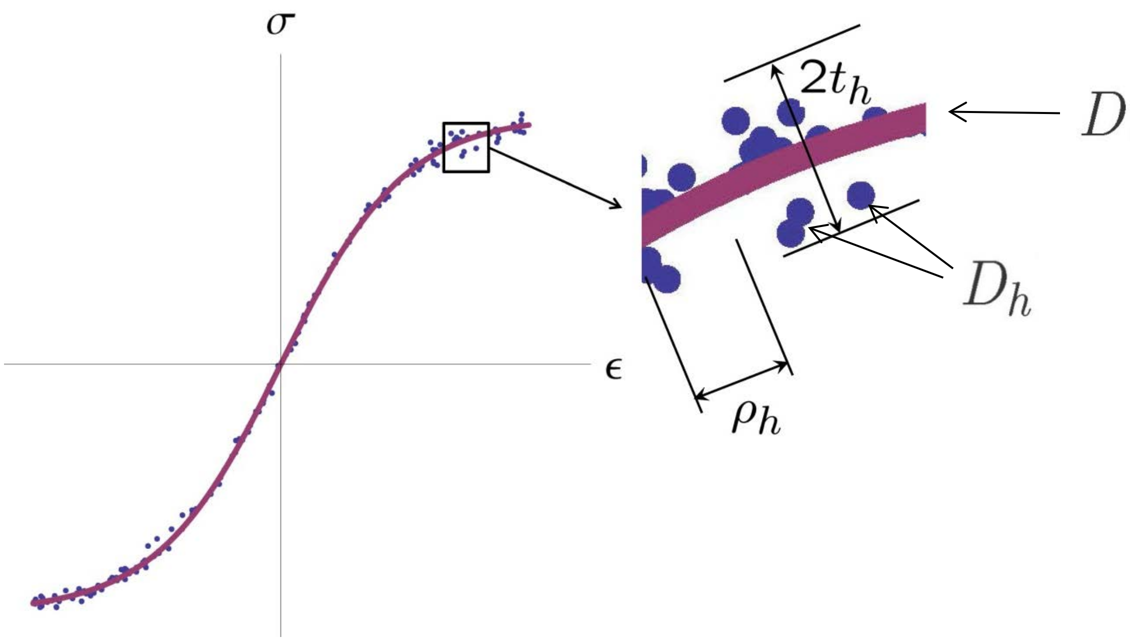
has a unique solution. Moreover, the Data-Driven solution satisfies

$$\sigma = \mathbb{C}\epsilon.$$

Remark: Theorem extends to general monotone functions

S. Conti, S. Müller and M. Ortiz (2017) arXiv:1708:02880

Data-Driven elasticity – Convergence with respect to the data set



Theorem

Suppose D *monotone graph,* $\rho_h \downarrow 0$ *and* $t_h \downarrow 0$ *such that:*

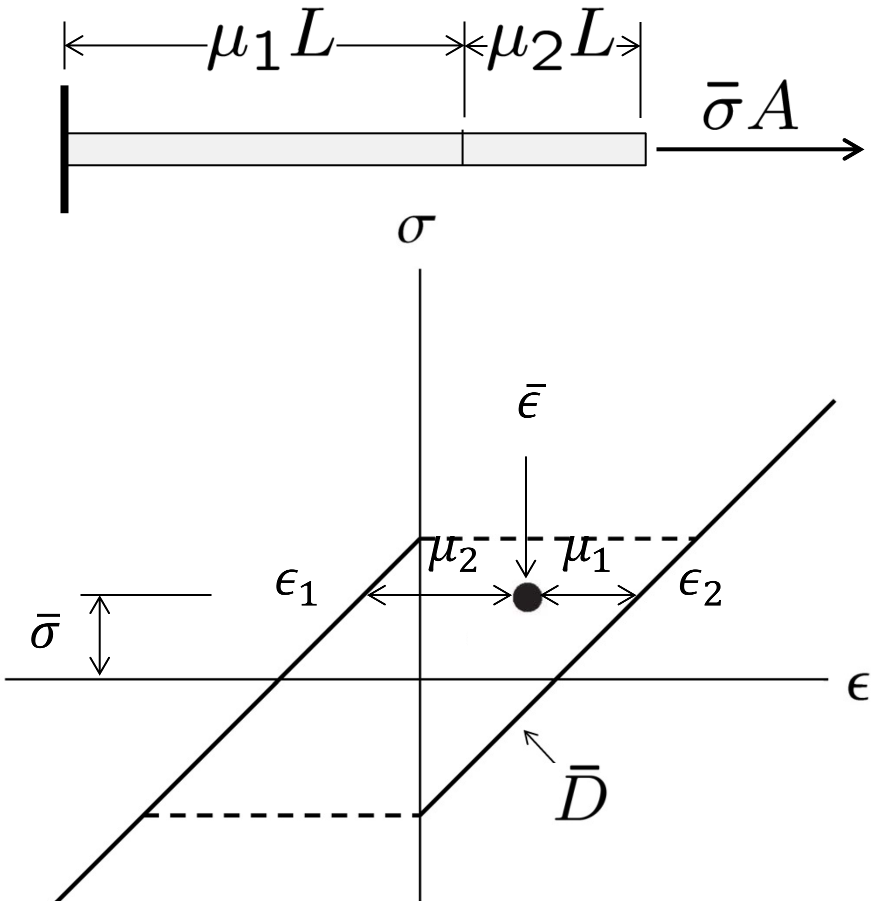
i) *Fine approximation:* $d(\xi, D_h) \leq \rho_h, \forall \xi \in D$.

ii) *Uniform approximation:* $d(\xi, D) \leq t_h, \forall \xi \in D_h$.

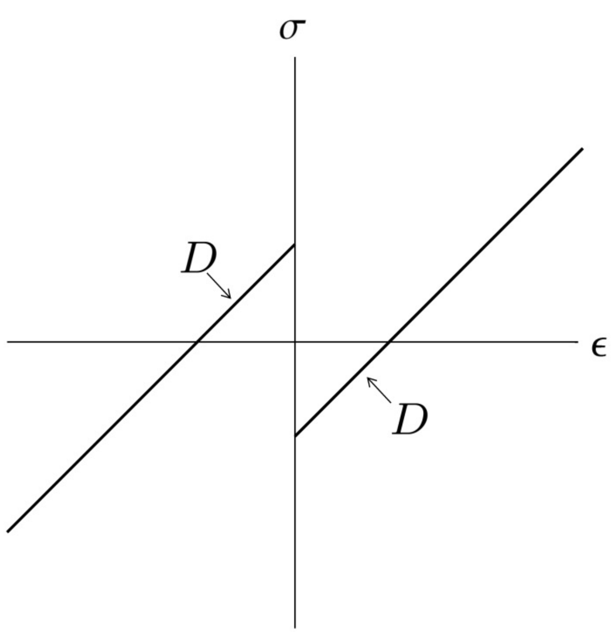
Then, $(\epsilon_h, \sigma_h) \rightarrow (\epsilon, \sigma)$.

S. Conti, S. Müller and M. Ortiz (2017) arXiv:1708:02880

Data-Driven elasticity - Relaxation



- Constraint set:
  - i) $\sigma(x) = \bar{\sigma} A.$
  - ii) $\bar{\epsilon} = \fint \epsilon(x)\, dx.$
- Data set (double well):



Theorem (DD Relaxation)

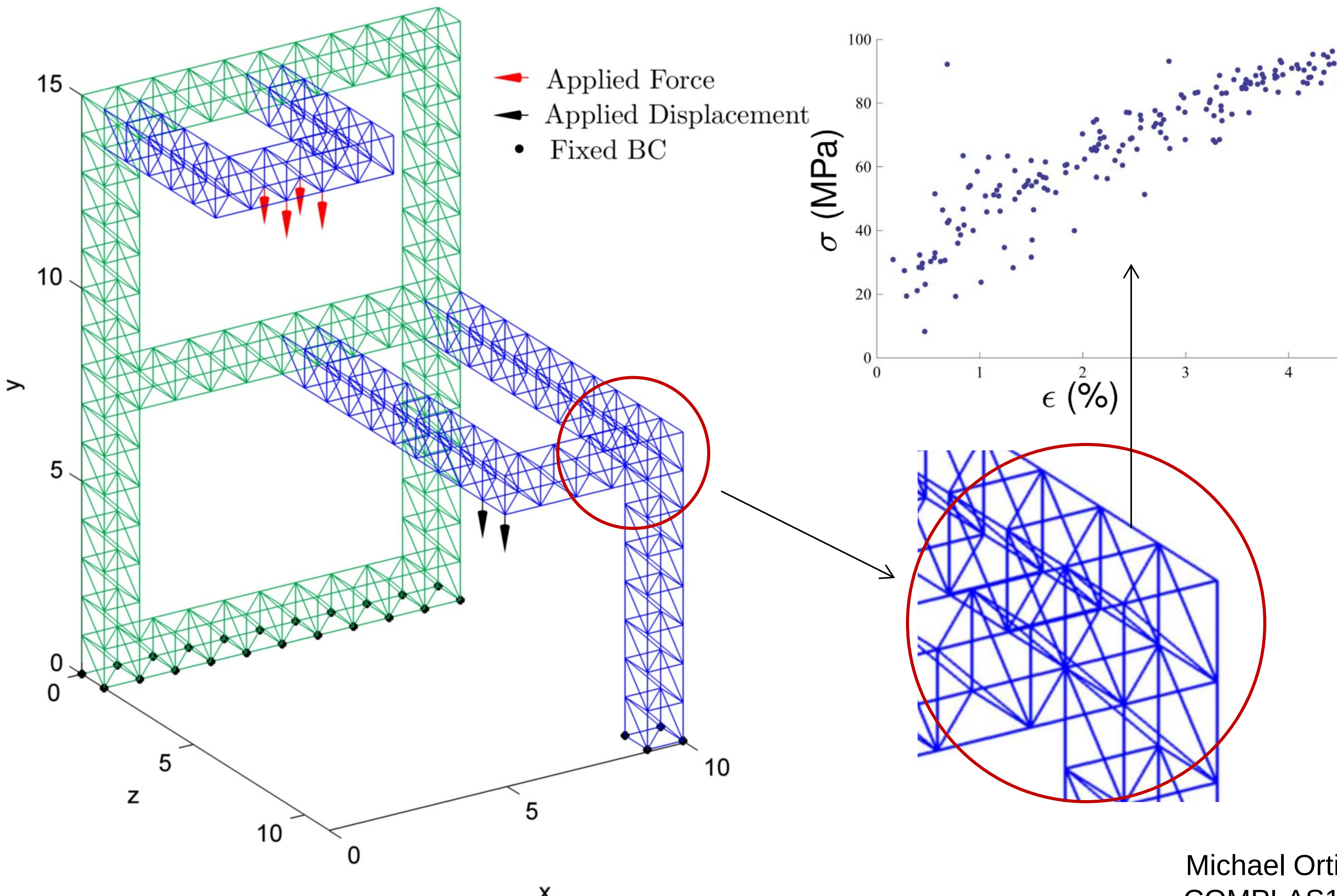
$$D \equiv \{\textit{double well}\} \Rightarrow \{(\bar{\epsilon}, \bar{\sigma})\} = \overline{D}.$$

S. Conti, S. Müller and M. Ortiz (2017) arXiv:1708:02880

Data-Driven Problems

- Data-driven problems represent a *complete reformulation* of the classical problems of mechanics (data + differential constraints)
- Data-driven problems *subsume*—and are strictly larger than—classical problems
- Data-driven analysis leads to notions of *convergence of data sets* that imply convergence of solutions.
- Data-driven *relaxation* (micro-macro) is fundamentally different from classical relaxation of energy functions!

Implementation: Trusses



Michael Ortiz
COMPLAS17

Implementation: Trusses

- Degrees of freedom: $(u_i)_{i=1}^{n}$
- Phase space: $\{(\epsilon_e, \sigma_e)_{e=1}^{m}\} \equiv X$, with norm

$$|(\boldsymbol{\epsilon}, \boldsymbol{\sigma})| = \sum_{e=1}^{m} w_e \left(\mathbb{C}\epsilon_e^2 + \mathbb{C}^{-1}\sigma_e^2 \right)$$

- Constraint set: $E = \{\boldsymbol{\epsilon} = \boldsymbol{B}\boldsymbol{u},\ \boldsymbol{B}^T\boldsymbol{\sigma} = \boldsymbol{f}\}$
- Data-driven problem:

$$\min_{(\boldsymbol{\epsilon}^*, \boldsymbol{\sigma}^*) \in D} \underbrace{\left(\min_{(\boldsymbol{\epsilon}, \boldsymbol{\sigma}) \in E} |(\boldsymbol{\epsilon} - \boldsymbol{\epsilon}^*, \boldsymbol{\sigma} - \boldsymbol{\sigma}^*)|^2 \right)}_{\text{Standard linear truss problem}}$$

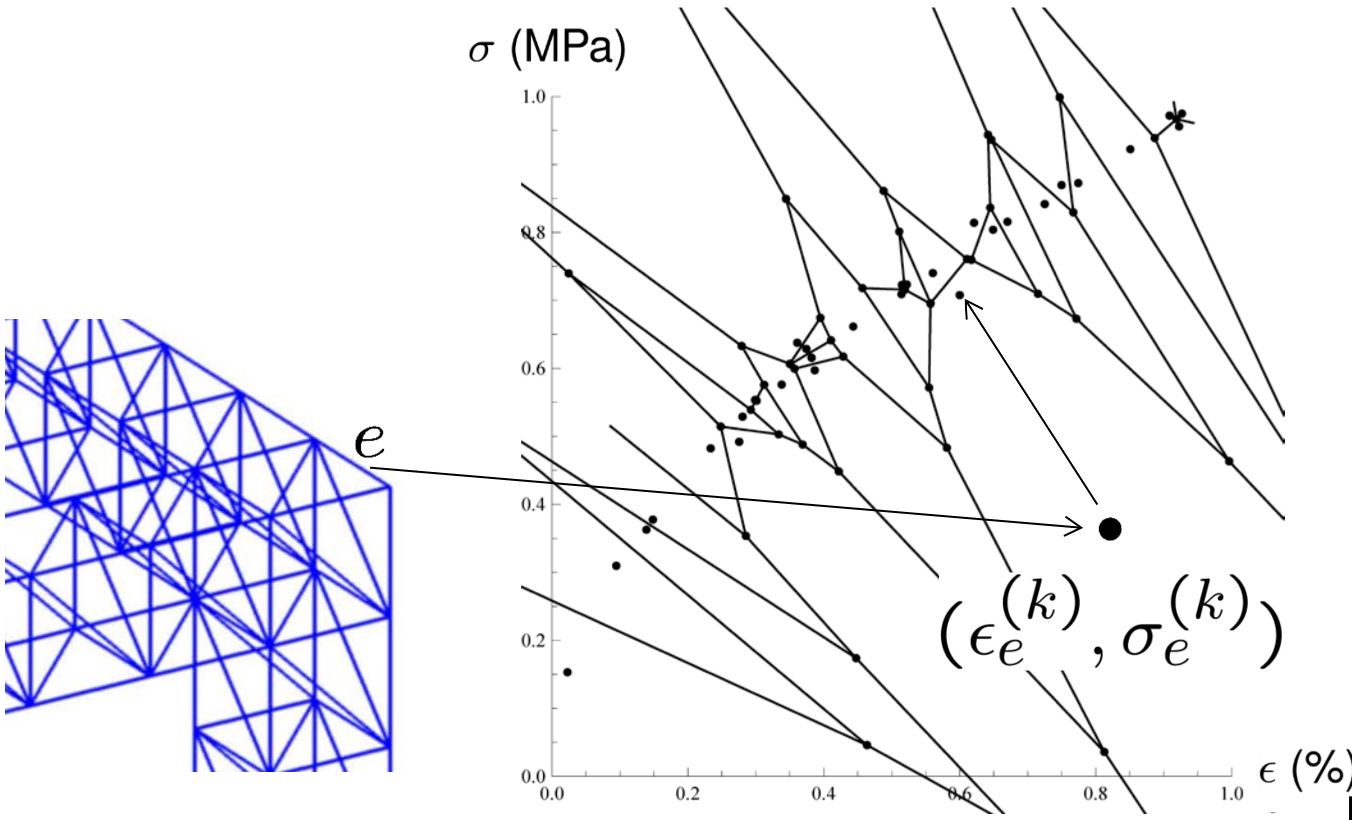
Standard linear truss problem

Search over material-data set

Implementation: Staggered solver

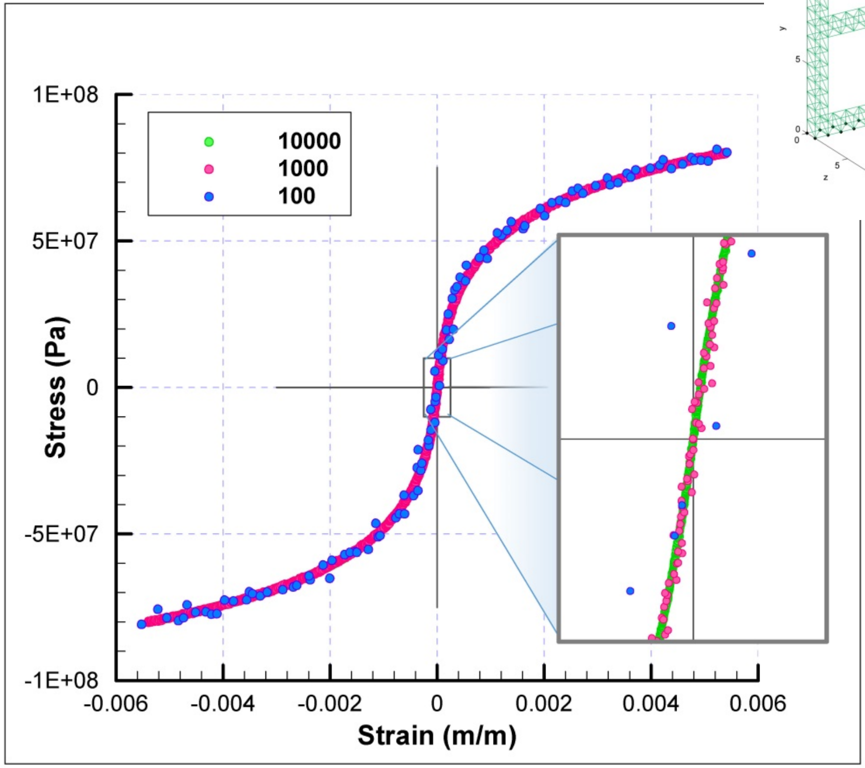
i) Mechanical step: $\min_{(\boldsymbol{\epsilon},\boldsymbol{\sigma})\in E} |(\boldsymbol{\epsilon}-\boldsymbol{\epsilon}^{*(k)},\boldsymbol{\sigma}-\boldsymbol{\sigma}^{*(k)})|^2$

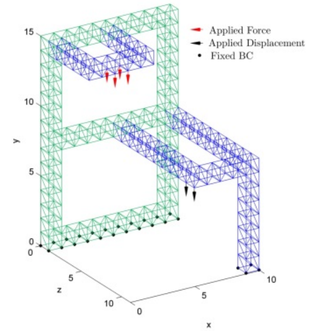
ii) Data association step:

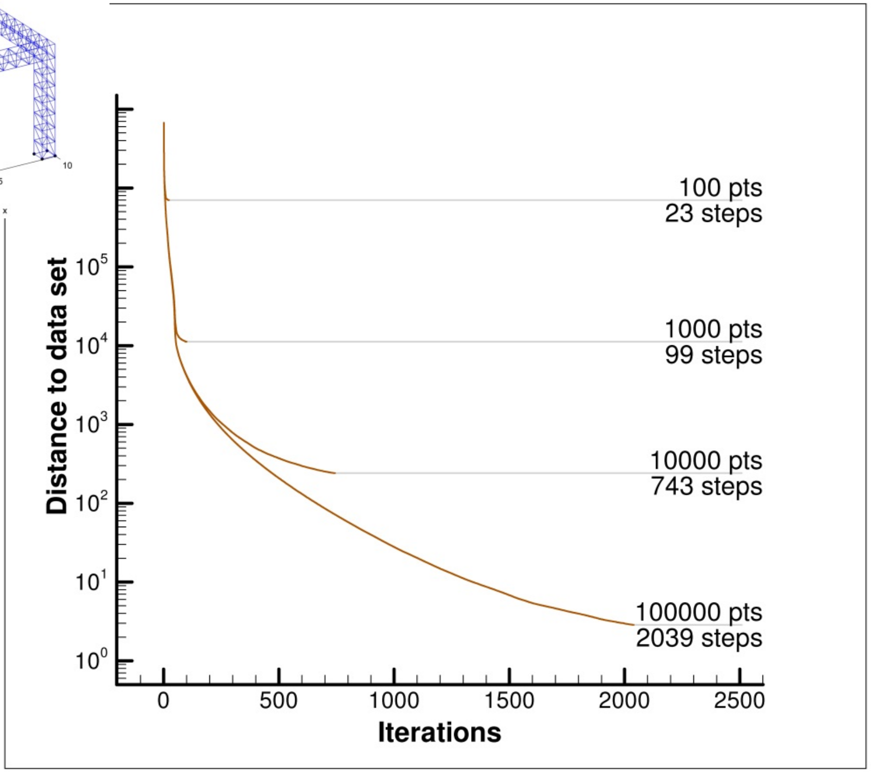


Michael Ortiz
COMPLAS17

Truss example: Convergence of solver





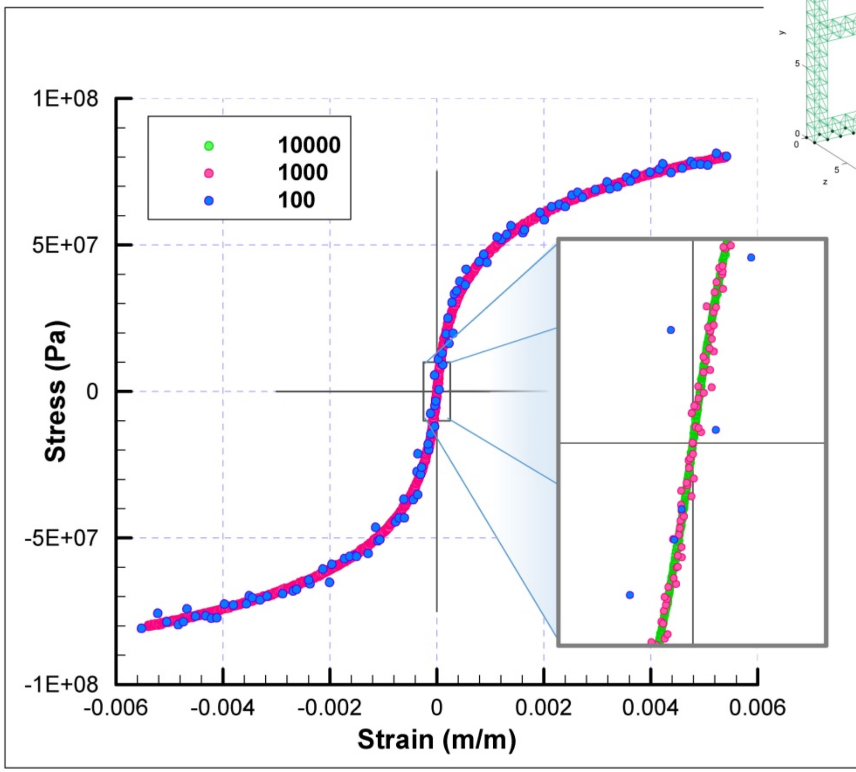


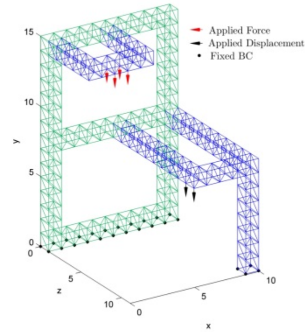
Material-data sets
of increasing size
and decreasing scatter

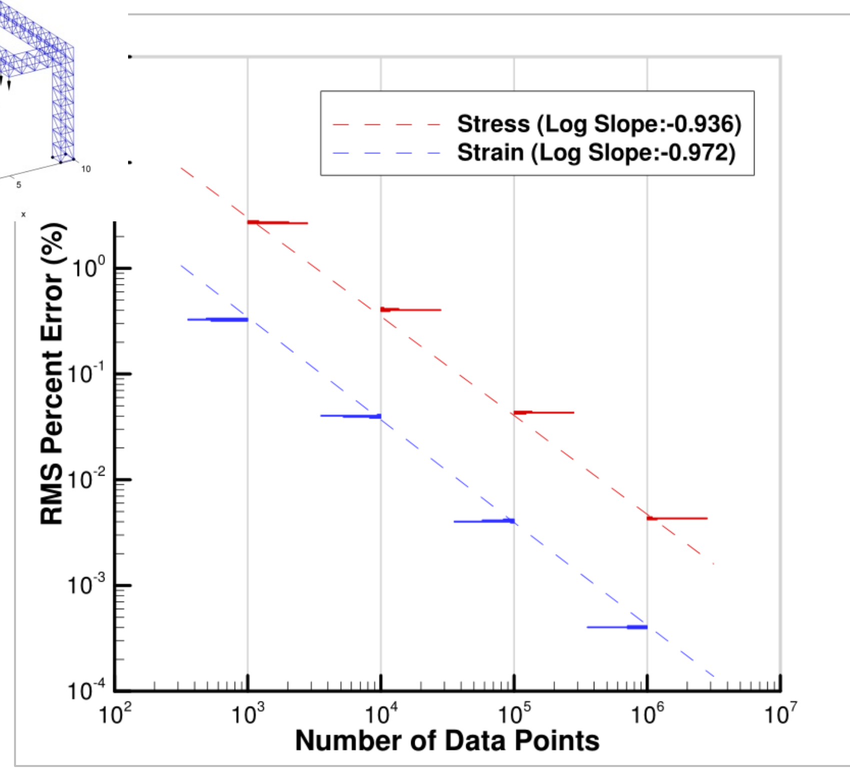
Convergence,
local data assignment
iteration

Michael Ortiz
COMPLAS17

Truss example: Convergence wrt data







Material-data sets
of increasing size
and decreasing scatter

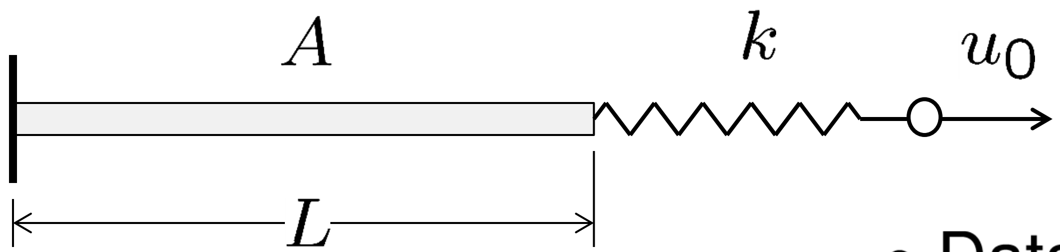
Convergence
with respect to sample size
(with initial Gaussian noise)

Michael Ortiz
COMPLAS17

Distance-based DD solvers

- Distance-based DD solvers exhibit good convergence wrt to material data associations
- Distance-based DD solvers exhibit good convergence wrt uniformly converging data
- But distance-based DD solvers can be overly sensitive to *outliers* in the data (non-uniform data convergence)
- If outliers cannot be ruled out, distance-based DD solvers need to be generalized and extended...

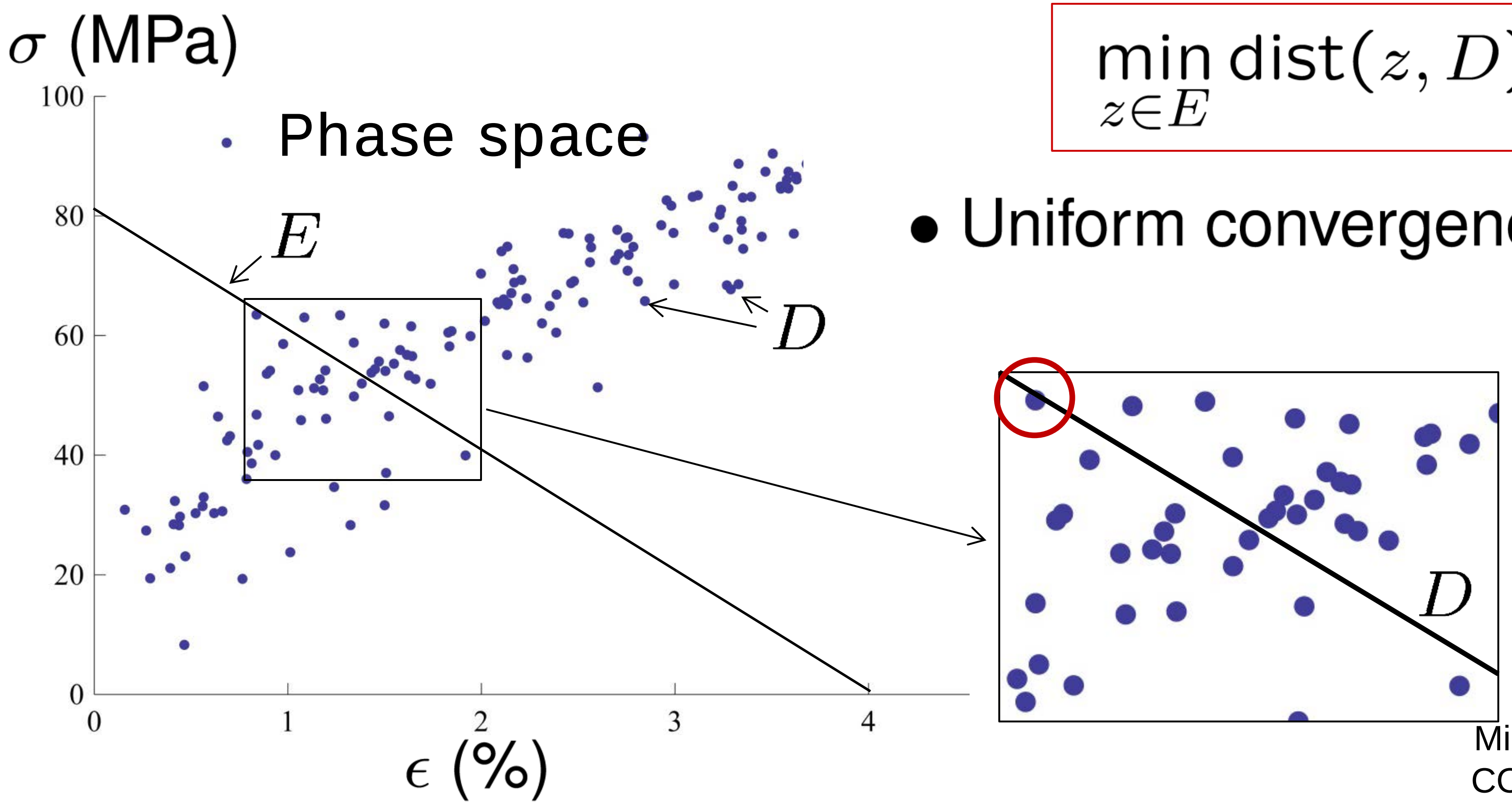
Extension to noisy data sets



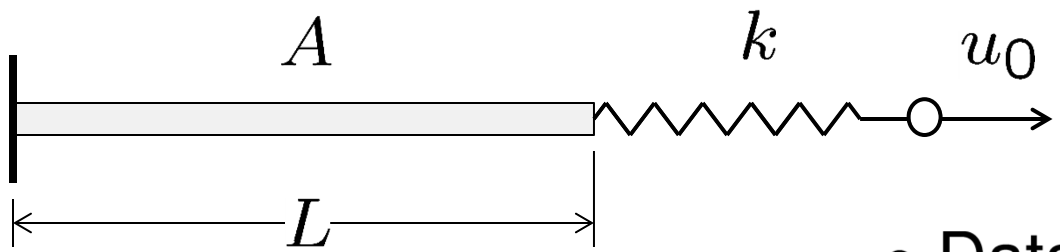
- Data-driven solution:

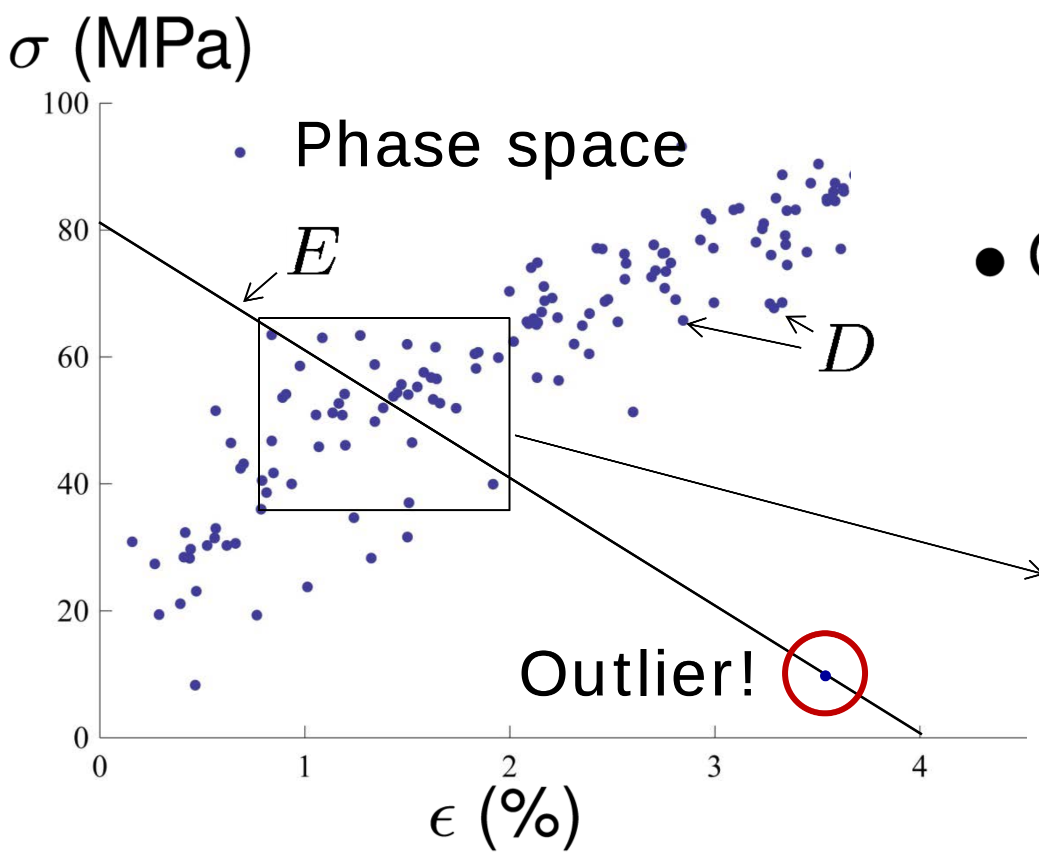
$$\min_{z \in E} \mathrm{dist}(z, D)$$

- Uniform convergence ✓



Extension to noisy data sets

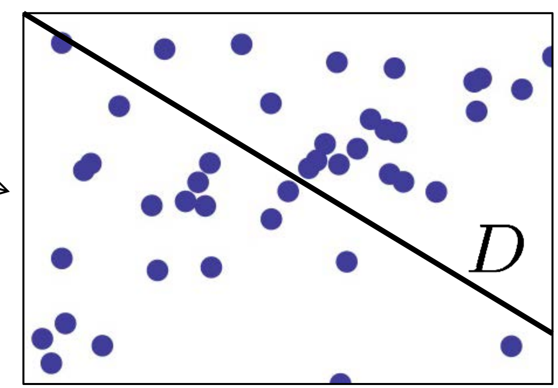




- Data-driven solution:

$$\min_{z \in E} \operatorname{dist}(z, D)$$

- Outliers may dominate!



Extension to noisy data sets

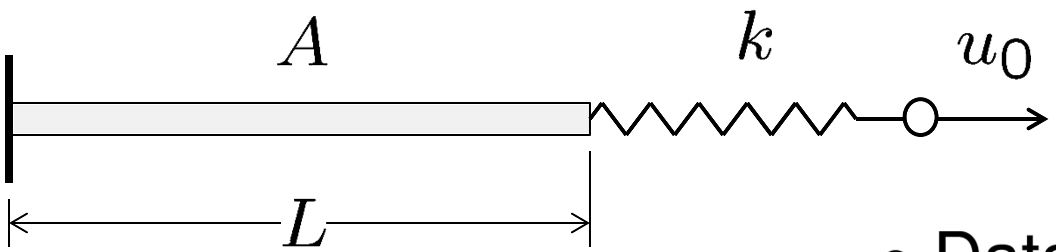
- Distance-based DD suffers from a *tyranny of the outliers* (non-uniform convergence)
- Eliminate by 'polling' the data set more widely (cluster analysis, *max-ent* inference...)
- '*Thermalize*' distance to material set $D = (z_1, \ldots, z_N)$:

$$d_\beta(z, D) = -\frac{1}{\beta}\log\left(\sum_{i=1}^{N} \mathrm{e}^{-(\beta/2)d^2(z,z_i)}\right)$$

- *Max-ent DD problem*[1]: $\mathrm{argmin}\{d_\beta(z, D),\ z \in E\}$
- Solve by *simulated annealing*!

[1] T. Kirchdoerfer and M. Ortiz (2017) arXiv:1702.01574v2

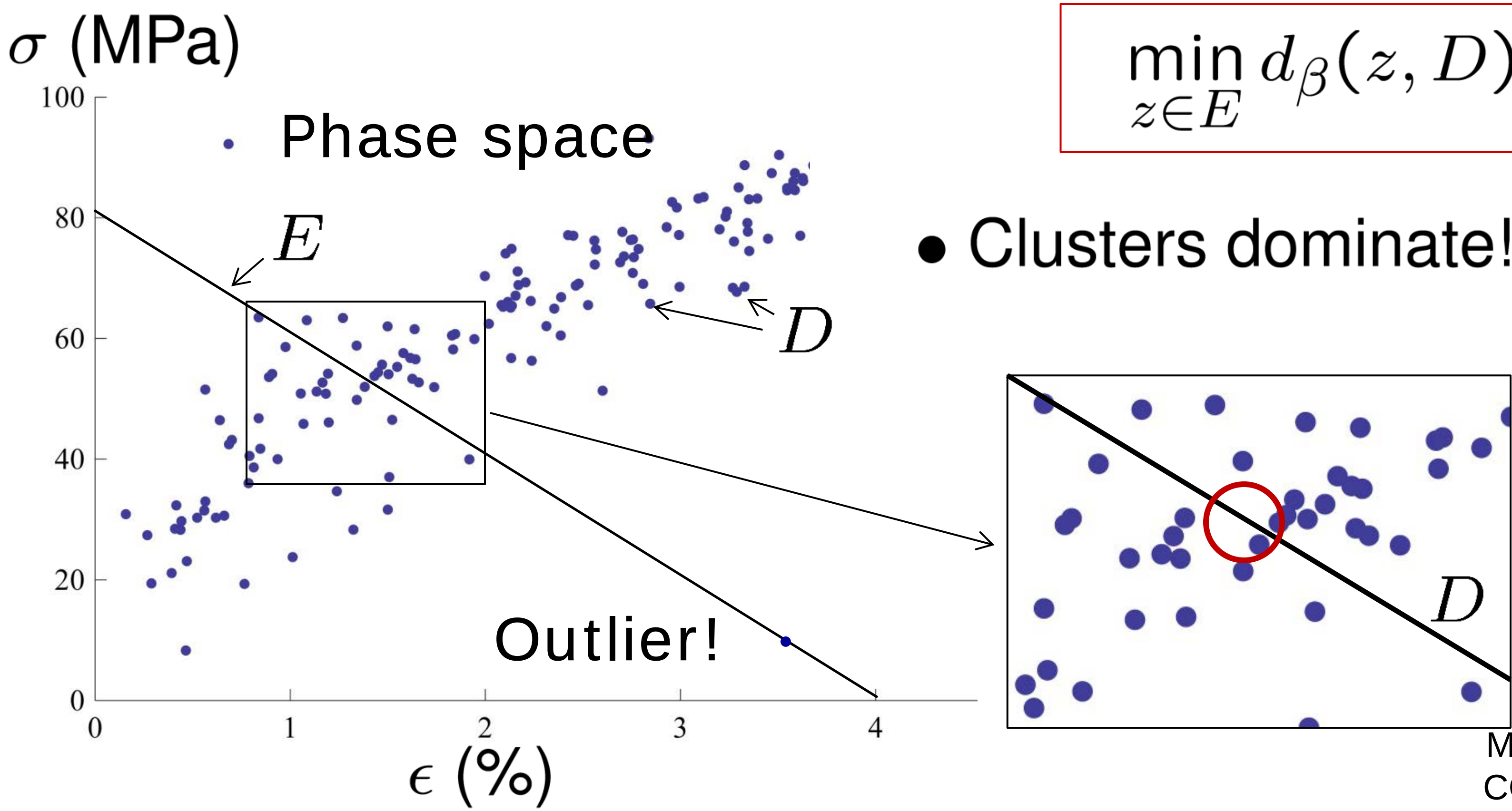
Extension to noisy data sets



- Data-driven solution:

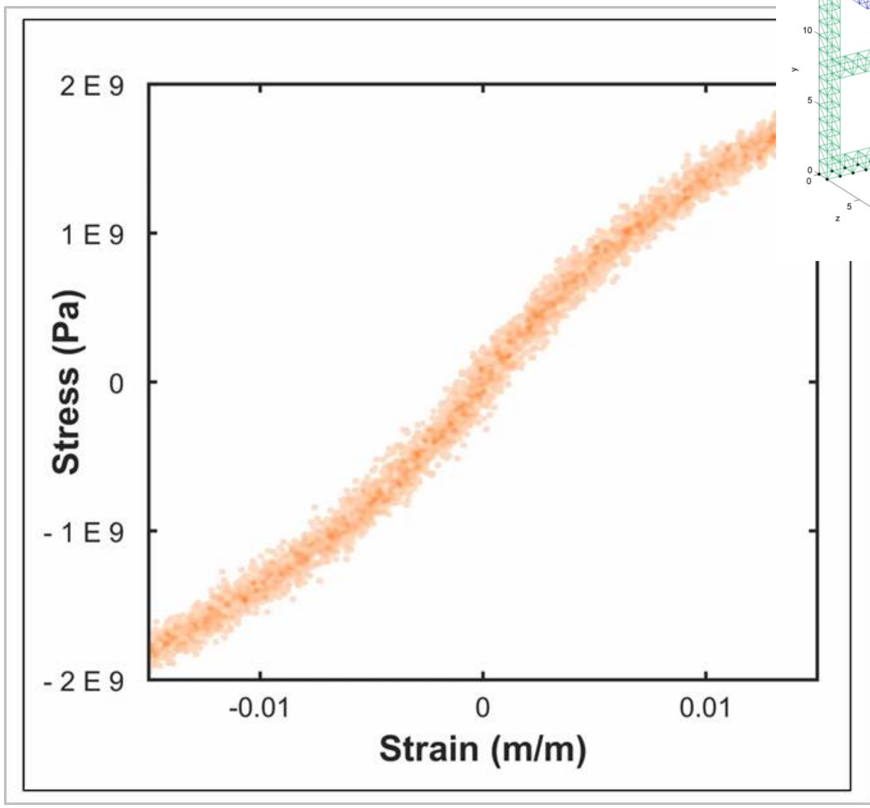
$$\min_{z \in E} d_\beta(z, D)$$

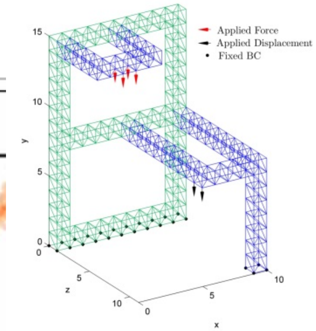
- Clusters dominate!

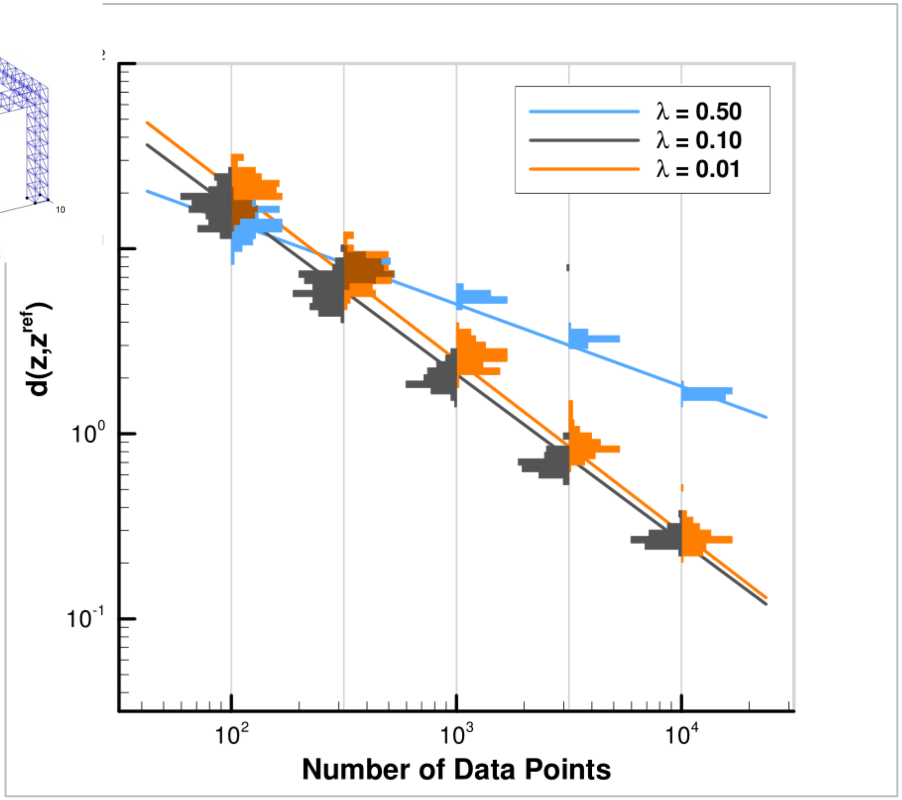


Michael Ortiz
COMPLAS17

Truss example: Convergence wrt data







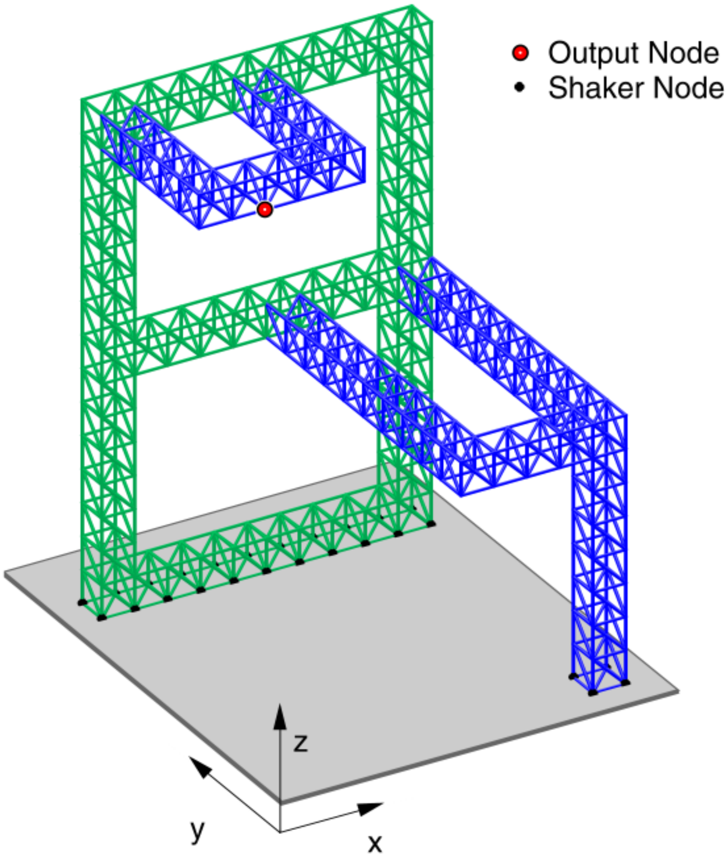
Material-data sets
of increasing size
and decreasing scatter

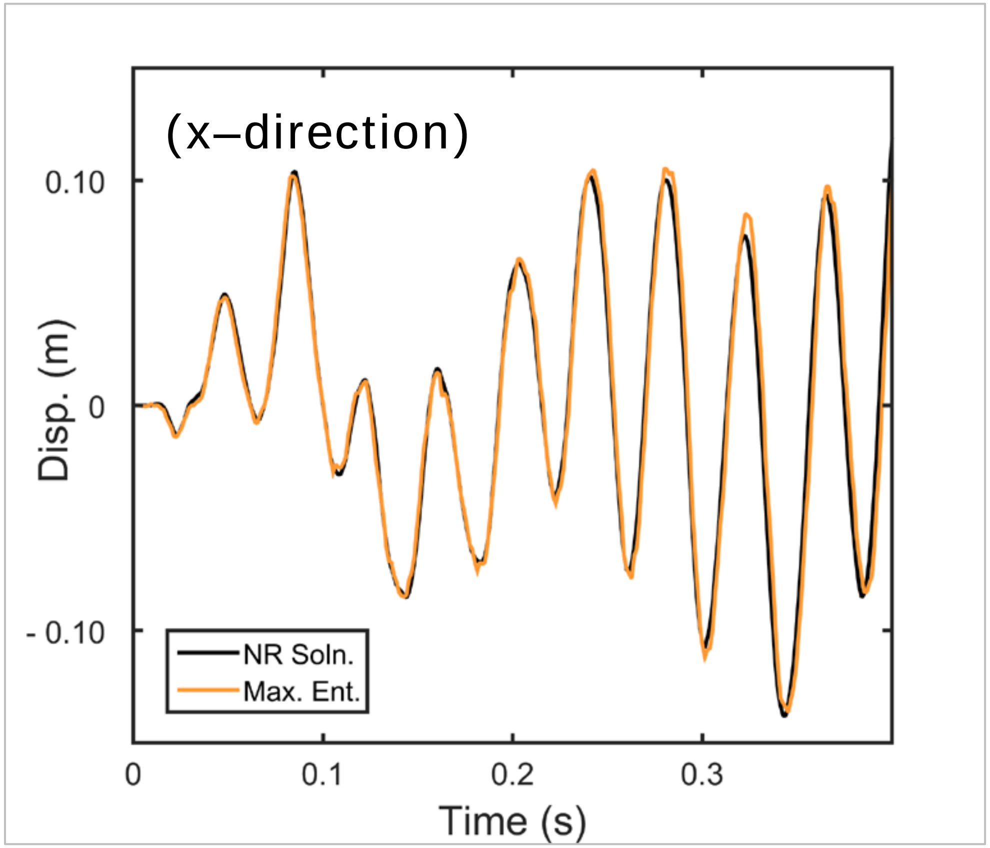
Convergence
with respect to sample size
(with Gaussian noise)

Michael Ortiz
COMPLAS17

Extension to dynamics

- Constraint set: Time-discrete eqs. of motion





[1]T. Kirchdoerfer and M. Ortiz (2017) arXiv:1702.01574v2

Michael Ortiz
COMPLAS17

Concluding remarks

- *Data-driven computing* is emerging as an alternative paradigm to model-based computing
- Data-driven computing can reliably supply solutions from raw material data sets
- Data-driven computing is likely to be a growth area in an increasingly *data-rich world*
- Numerous outstanding questions:
  - *Phase-space coverage, importance sampling*
  - *Building goal-oriented material data bases from experiment[1] and from first-principles calculations*
  - *Inelasticity, path dependence...*

[1]A. Leygue *et al*. (2017) HAL Id: hal-01452494.

Concluding remarks

Michael Ortiz
COMPLAS17