Linear Multistep Methods for Learning Dynamics

Qiang Du

Department of Applied Physics and Applied Mathematics,
and Data Science Institute, Columbia University

Joint work with **Rachael Keller (Columbia)**; **Yiqi Gu** and **Chao Zhou** (NUS), **Haizhao Yang** (Purdue)

Thanks to INAF organizers for the invitation

Feb 2022 Irish Numerical Analysis Forum

About this lecture

Question:

Is a good numerical scheme for integrating known dynamics also good for learning unknown dynamics?

$\Rightarrow$ New convergence theory on Linear Multistep Method (LMM) for dynamics discovery.

References: Keller-Du, Discovery of dynamics using linear multistep methods,
SIAM J. Numer. Anal., 59, 429–455, 2021 (arXiv 1912.12728).

and a follow-up work: Du-Gu-Yang-Zhou, arXiv:2103.11488, 2021.

Classical Linear Multistep Method (LMM)

LMM[1]: numerical integrator for $\dot{x}(t)=f(x(t))$, $t\in(0,T)$, $x(0)=x_0$ with given f, x_0, T and $M\in\mathbb{N}$.

For $N\in\mathbb{N}$, a uniform step size $h=T/N$, and approximate initial states $\{\boldsymbol{x}_n \sim x(nh)\}_{n=0}^{M-1}$,

$$\boxed{\sum_{m=0}^{M} \alpha_m \boldsymbol{x}_{n-m} = h\sum_{m=0}^{M} \beta_m f(\boldsymbol{x}_{n-m}), \quad M\le n\le N} \Rightarrow \quad \text{approximate states } \{\boldsymbol{x}_n \sim x(nh)\}_{n=M}^{N}.$$

[1] LMM: a subject taught in a typical course on the numerical solution of differential equations.

Classical Linear Multistep Method (LMM)

LMM[1]: numerical integrator for $\dot{x}(t)=f(x(t))$, $t\in(0,T)$, $x(0)=x_0$ with given f, x_0, T and $M\in\mathbb{N}$.

For $N\in\mathbb{N}$, a uniform step size $h=T/N$, and approximate initial states $\{\boldsymbol{x}_n \sim x(nh)\}_{n=0}^{M-1}$,

$$\boxed{\sum_{m=0}^{M}\alpha_m \boldsymbol{x}_{n-m} = h\sum_{m=0}^{M}\beta_m f(\boldsymbol{x}_{n-m}),\quad M\le n\le N} \Rightarrow \quad \text{approximate states } \{\boldsymbol{x}_n \sim x(nh)\}_{n=M}^{N}.$$

Many successful applications and well-established mathematical theory for LMM since 1883[2].

Eg. Adams-Bashforth, Adams-Moulton, Backward-Differentiation-Formula (BDF).

Key ingredients: 1st/2nd characteristic polynomials, $\rho(r)=\sum_{m=0}^{M}\alpha_m r^{M-m}$, $\sigma(r)=\sum_{m=0}^{M}\beta_m r^{M-m}$.

[2] Atkinson-Han-Stewart 11, Butcher 03, Gautschi 97, Hairer-Norsett-Wanner 93, Henrici 62, Iserles 96, Mayers-Euli 03, etc.

Classical Linear Multistep Method (LMM)

LMM[1]: numerical integrator for $\dot{x}(t)=f(x(t))$, $t\in(0,T)$, $x(0)=x_0$ with given f, x_0, T and $M\in\mathbb{N}$.

For $N\in\mathbb{N}$, a uniform step size $h=T/N$, and approximate initial states $\{\boldsymbol{x}_n \sim x(nh)\}_{n=0}^{M-1}$,

$$\boxed{\sum_{m=0}^{M}\alpha_m \boldsymbol{x}_{n-m} = h\sum_{m=0}^{M}\beta_m f(\boldsymbol{x}_{n-m}),\quad M\le n\le N} \Rightarrow \text{ approximate states } \{\boldsymbol{x}_n \sim x(nh)\}_{n=M}^{N}.$$

Many successful applications and well-established mathematical theory for LMM since 1883[2].

Eg. Adams-Bashforth, Adams-Moulton, Backward-Differentiation-Formula (BDF).

Key ingredients: 1st/2nd characteristic polynomials, $\rho(r)=\sum_{m=0}^{M}\alpha_m r^{M-m}$, $\sigma(r)=\sum_{m=0}^{M}\beta_m r^{M-m}$.

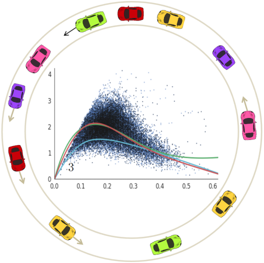
$$\left.\begin{array}{l}\text{Consistency: } \rho(1)=0,\ \rho'(1)=\sigma(1).\\ \text{Stability: } \rho(r)=0 \Leftrightarrow |r|\le 1, \text{ simple if } |r|=1.\end{array}\right\} \Rightarrow \text{ Convergence (classical theory)}^{3}$$

[3] Dahlquist 1955, Henrici 1962,...: turning convergence of LMM for integrating dynamics into pure algebraic criteria

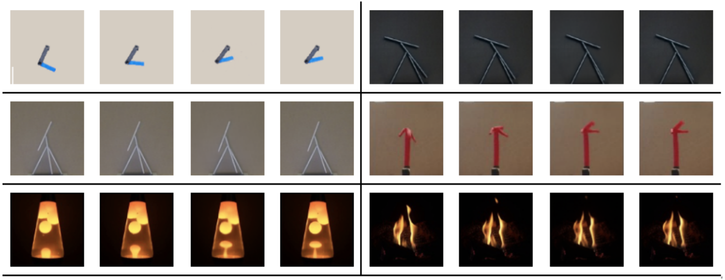
Data-driven learning and modeling via machine learning

Motivations: data-driven modeling via deep learning (an inverse problem). Examples:

- Learning traffic model from data (partnership with COSMOS, funded by AWS, NSF and FHWA);
- Video-based learning (joint project with robotics lab at Columbia).



R. Shi, K. Huang, Z. Mo, X. Di, Q. Du,
IEEE Trac ITS 2021



B. Chen, K. Huang, S. Raghupathi, I. Chandratreya, Q. Du, H. Lipson, 2021

A new chapter on LMM[4]

Data-driven modeling via deep learning (an inverse problem) $\Rightarrow$ new applications of LMM[5].

[4] Kelller-Du SINUM 2021; Du-Gu-Yang-Zhou, arXiv:2103.11488.

[5] For example, discretizing NeuralODE/ResNet: E 2017, Chen-Rubanova-Bettencourt-Duvenaud, NIPS 2018, Lu-Zhong-Li-Dong ICML 2018,...

A new chapter on LMM[4]

Data-driven modeling via deep learning (an inverse problem) $\Rightarrow$ new applications of LMM[5].

Forward Prob.: Integrating dynamics

given $\dot{x}(t) = f(x(t))$ (physics),

with $x(0) = x_0$, solve for $x(t)$ (data).

Inverse Prob.: Learning dynamics

given observed $\{\boldsymbol{x}_n\}$ of $x(t)$ (data),

learn dynamics $\dot{x}(t) = f(x(t))$ (physics).

[4] Kellller-Du SINUM 2021; Du-Gu-Yang-Zhou, arXiv:2103.11488.

[5] For example, discretizing NeuralODE/ResNet: E 2017, Chen-Rubanova-Bettencourt-Duvenaud, NIPS 2018, Lu-Zhong-Li-Dong ICML 2018,...

A new chapter on LMM[4]

Data-driven modeling via deep learning (an inverse problem) $\Rightarrow$ new applications of LMM[5].

Forward Prob.: Integrating dynamics	**Inverse Prob.**: Learning dynamics
given $\dot{x}(t)=f(x(t))$ (physics),	given observed $\{\boldsymbol{x}_n\}$ of $x(t)$ (data),
with $x(0)=x_0$, solve for $x(t)$ (data).	learn dynamics $\dot{x}(t)=f(x(t))$ (physics).

Close connection to numerical ODE/PDE, model reduction, closure and multiscale simulations[6]:

Given dynamics $\dot{x}=f(x)$ $\Rightarrow$ solve for snapshots $\{\boldsymbol{x}_n\}$ $\Rightarrow$ effective dynamics $\dot{x}=\hat{f}(x)$

fine scale models $\rightarrow$ numerical solution + data $\rightarrow$ coarse scale models

Forward Problem — Inverse Problem

[6] Nadler-Lafon-Coifman-Kevrekidis 06, Khoo-Lu-Ying 17, Cao-Wang-E 18, Zhu-Zabaras 18, Bhattacharya-Hosseini-Kovachki-Stuart, 18, Du-Gunzburger 02, Qian-Kramer-Peherstorfer-Willcox 20, Lee-Carlberg 20, Wang-Cheung-Leung-Chung-Efendiev-Wheeler 20,......

The Era of Informative and Intelligent Scientific Computation (I^2SC)

⋆ (I^2SC) represents the loop of discovery: data-driven modeling/learning + integrating dynamics.

Informative and Intelligent

Scientific Computation (I^2SC)

Model

Physics

Data

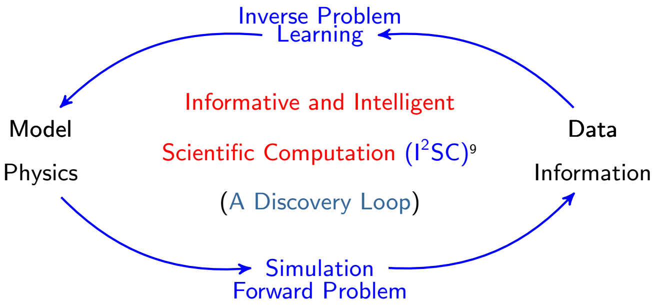
Information

Simulation
Forward Problem

[9] I^2SC was coined in 2004, published later in Du 2008, AIP Studies in Applied Math

The Era of Informative and Intelligent Scientific Computation (I^2SC)

⋆ (I^2SC) represents the loop of discovery: data-driven modeling/learning + integrating dynamics.



[9] I^2SC was coined in 2004, published later in Du 2008, AIP Studies in Applied Math

Data-driven modeling of dynamics via machine learning

Data-driven machine learning of $\dot{x} = f(x)$: find $f = f(x)$ from given snapshots $\{\boldsymbol{x}_n\}$ of state $x(t)$.

Some relevant works: Schmidt-Lipson, *Distilling free-form natural laws from experimental data*, Science, 09
Brunton-Noack-Koumoutsakos ARFM 20,
Kang-Liao-Liu JSC 21,
Kevrekidis- Rowley-Williams JCD 16,
Kutz-Brunton-Brunton-Proctor SIAM 16,
Lee-Carlberg, JCP 20,
Long-Lu-Dong JCP 19,
Lu-Zhong-Tang-Maggioni PNAS 19,
Qian-Kramer-Peherstorfer-Willcox PhysD 20,
Qin-Wu-Xiu JCP 19,
Raissi-Perdikaris-Karniadakis 18,
Raissi-Yazdani-Karniadakis Science 20,
Rudy-Kutz-Brunton JCP 19,
Sun-Zhang-Schaeffer PMLR 19,
Teng-Wang-Ding-Zhang-Wang 20,
Tipireddy-Perdikaris-Stinis-Tartakovsky 19,
Wang-Cheung-Leung-Chung-Efendiev-Wheeler 20,
Wu-Xiu JCP 20,
Xie-Zhang-Webster 18,
......

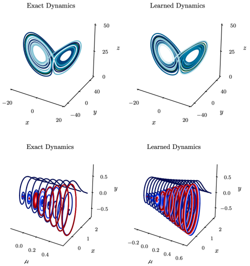
Data-driven modeling of dynamics via machine learning

Data-driven machine learning of $\dot{x} = f(x)$: find $f = f(x)$ from given snapshots $\{\boldsymbol{x}_n\}$ of state $x(t)$.

Some relevant works: Schmidt-Lipson, *Distilling free-form natural laws from experimental data*, Science, 09

Brunton-Noack-Koumoutsakos ARFM 20,
Kang-Liao-Liu JSC 21,
Kevrekidis- Rowley-Williams JCD 16,
Kutz-Brunton-Brunton-Proctor SIAM 16,
Lee-Carlberg, JCP 20,
Long-Lu-Dong JCP 19,
Lu-Zhong-Tang-Maggioni PNAS 19,
Qian-Kramer-Peherstorfer-Willcox PhysD 20,
Qin-Wu-Xiu JCP 19,
Raissi-Perdikaris-Karniadakis 18,
Raissi-Yazdani-Karniadakis Science 20,
Rudy-Kutz-Brunton JCP 19,
Sun-Zhang-Schaeffer PMLR 19,
Teng-Wang-Ding-Zhang-Wang 20,
Tipireddy-Perdikaris-Stinis-Tartakovsky 19,
Wang-Cheung-Leung-Chung-Efendiev-Wheeler 20,
Wu-Xiu JCP 20,
Xie-Zhang-Webster 18,
......

An illustrative attempt (LMnet):



DNN for f
+
LMM for dynamics.

R-P-K 18, ...

As expected?

Convergence?

Quantitative analysis/theorems?

Dynamics Discovery Using LMM: Numerical Analysis

Dynamics discovery of $\dot{x}(t) = f(x(t))$: given $\hat{\boldsymbol{x}}^h = \{\hat{\boldsymbol{x}}_n\}_{n=0}^N \Rightarrow LMM + DNN \Rightarrow f$.

Objective: to conform with (discrete) dynamics, and to achieve fidelity with data.

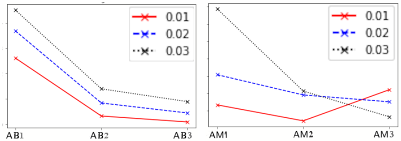
Expected: higher order LMM + more data + larger NN $\Rightarrow$ better recovered dynamics.

Dynamics Discovery Using LMM: Numerical Analysis

Dynamics discovery of $\dot{x}(t) = f(x(t))$: given $\hat{\boldsymbol{x}}^h = \{\hat{\boldsymbol{x}}_n\}_{n=0}^N \Rightarrow LMM + DNN \Rightarrow f$.

Objective: to conform with (discrete) dynamics, and to achieve fidelity with data.

Expected: higher order LMM + more data + larger NN $\Rightarrow$ better recovered dynamics.



Surprise! E.g., errors for a 2D harmonic oscillator: with Adams-Bashforth (AB, left) or Adams-Moulton (AM, right), NN with 256 nodes, tanh activation, h=.03, .02, .01.

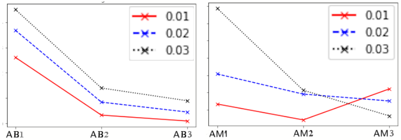
AM3: higher order, more data $\Rightarrow$ larger error.

Dynamics Discovery Using LMM: Numerical Analysis

Dynamics discovery of $\dot{x}(t) = f(x(t))$: given $\hat{\boldsymbol{x}}^h = \{\hat{\boldsymbol{x}}_n\}_{n=0}^N \Rightarrow LMM + DNN \Rightarrow f$.

Objective: to conform with (discrete) dynamics, and to achieve fidelity with data.

Expected: higher order LMM + more data + larger NN $\Rightarrow$ better recovered dynamics.



Surprise! E.g., errors for a 2D harmonic oscillator: with Adams-Bashforth (AB, left) or Adams-Moulton (AM, right), NN with 256 nodes, tanh activation, h=.03, .02, .01.

AM3: higher order, more data $\Rightarrow$ larger error.

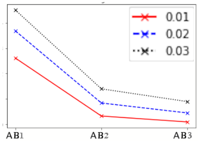
Explaining the unexpected: problems with training? landscape?...... the usual cliche.

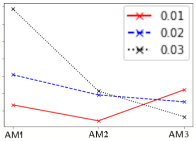
Dynamics Discovery Using LMM: Numerical Analysis

Dynamics discovery of $\dot{x}(t) = f(x(t))$: given $\hat{\boldsymbol{x}}^h = \{\hat{\boldsymbol{x}}_n\}_{n=0}^N \Rightarrow LMM + DNN \Rightarrow f$.

Objective: to conform with (discrete) dynamics, and to achieve fidelity with data.

Expected: higher order LMM + more data + larger NN $\Rightarrow$ better recovered dynamics.





Surprise! E.g., errors for a 2D harmonic oscillator: with Adams-Bashforth (AB, left) or Adams-Moulton (AM, right), NN with 256 nodes, tanh activation, h=.03, .02, .01.

AM3: higher order, more data $\Rightarrow$ larger error.

Explaining the unexpected: problems with training? landscape?...... the usual cliche.

Deeper reason: new tale of LMM, motivated by deep learning of unknown dynamics

(Kelller-Du arXiv:1912.12728, 2019,(SINUM 2021), Du-Gu-Yang-Zhou, arXiv2103.11488)

Dynamics Discovery Using LMM

Dynamics discovery of $\dot{x}(t) = f(x(t))$: given $\hat{\boldsymbol{x}}^h = \{\hat{x}_n\}_{n=0}^N \Rightarrow LMM + DNN \Rightarrow f$.

Framework: for estimated state/dynamics $(\boldsymbol{x}^h, \boldsymbol{f}^h)$ and NN represented dynamics $\hat{\boldsymbol{f}}_{NN}^h$ (discrete).

$$\boxed{\begin{aligned}\text{Loss} &:= \text{errors (dynamics conformity + NN representation + data fidelity) + regularization} \\ &= \gamma_1 \|\text{LMM residual of } (\boldsymbol{x}^h, \boldsymbol{f}^h)\|^2 + \gamma_2 \|\hat{\boldsymbol{f}}_{NN}^h - \boldsymbol{f}^h\|^2 + \gamma_3 \|\boldsymbol{x}^h - \hat{\boldsymbol{x}}^h\|^2 + \gamma_4 R(\hat{\boldsymbol{f}}_{NN}^h, \boldsymbol{f}^h, \boldsymbol{x}^h).\end{aligned}}$$

Important to note: only connection to the unknown dynamics is provided by the LMM residual (effectively linking the differentiation of the state with the interpolation of the dynamics).

Dynamics Discovery Using LMM

Dynamics discovery of $\dot{x}(t) = f(x(t))$: given $\hat{\boldsymbol{x}}^h = \{\hat{\boldsymbol{x}}_n\}_{n=0}^N \Rightarrow LMM + DNN \Rightarrow f$.

Framework: for estimated state/dynamics $(\boldsymbol{x}^h, \boldsymbol{f}^h)$ and NN represented dynamics $\hat{\boldsymbol{f}}_{NN}^h$ (discrete).

$$\boxed{\begin{aligned}\text{Loss} &:= \text{errors (dynamics conformity + NN representation + data fidelity) + regularization} \\ &= \gamma_1 \|\text{LMM residual of } (\boldsymbol{x}^h, \boldsymbol{f}^h)\|^2 + \gamma_2 \|\hat{\boldsymbol{f}}_{NN}^h - \boldsymbol{f}^h\|^2 + \gamma_3 \|\boldsymbol{x}^h - \hat{\boldsymbol{x}}^h\|^2 + \gamma_4 R(\hat{\boldsymbol{f}}_{NN}^h, \boldsymbol{f}^h, \boldsymbol{x}^h).\end{aligned}}$$

Important to note: only connection to the unknown dynamics is provided by the LMM residual (effectively linking the differentiation of the state with the interpolation of the dynamics).

Dynamics Discovery Using LMM

Dynamics discovery of $\dot{x}(t) = f(x(t))$: given $\hat{\boldsymbol{x}}^h = \{\hat{x}_n\}_{n=0}^N \Rightarrow LMM + DNN \Rightarrow f$.

Framework: for estimated state/dynamics $(\boldsymbol{x}^h, \boldsymbol{f}^h)$ and NN represented dynamics $\hat{\boldsymbol{f}}^h_{NN}$ (discrete).

$$\boxed{\begin{aligned} \text{Loss} &:= \text{errors (dynamics conformity + NN representation + data fidelity) + regularization} \\ &= \gamma_1 \|\text{LMM residual of } (\boldsymbol{x}^h, \boldsymbol{f}^h)\|^2 + \gamma_2 \|\hat{\boldsymbol{f}}^h_{NN} - \boldsymbol{f}^h\|^2 + \gamma_3 \|\boldsymbol{x}^h - \hat{\boldsymbol{x}}^h\|^2 + \gamma_4 R(\hat{\boldsymbol{f}}^h_{NN}, \boldsymbol{f}^h, \boldsymbol{x}^h). \end{aligned}}$$

Important to note: only connection to the unknown dynamics is provided by the LMM residual (effectively linking the differentiation of the state with the interpolation of the dynamics).

An idealized setting (Keller-Du 2019): unique/smooth dynamics in $[0, T]$; complete and exact snapshots of the state ($\hat{\boldsymbol{x}}^h = \boldsymbol{x}^h$); no regularization; error-free NN representation ($\hat{\boldsymbol{f}}^h_{NN} = \hat{\boldsymbol{f}}^h$).

$\Rightarrow$ Loss for dynamics discovery is primarily due to the LMM residual of $(\hat{\boldsymbol{x}}^h, \hat{\boldsymbol{f}}^h)$.

$\Rightarrow$ Minimize loss $\Leftrightarrow$ Zero LMM residual of $(\hat{\boldsymbol{x}}^h, \hat{\boldsymbol{f}}^h)$ (overparametrized regime).

New theory on convergence for learning $\boldsymbol{f}^h - \hat{\boldsymbol{f}}^h_{NN} = \boldsymbol{f}^h - \hat{\boldsymbol{f}}^h \to 0$ as $h \to 0$ via consistency/stability.

LMM: integrating and learning dynamics

Mathematical theory on LMM for integrating and learning dynamics		
Task	Integrating dynamics	Learning dynamics
Goal	Given $\boldsymbol{f}$, find $\boldsymbol{x}(t)$	Given $\{\boldsymbol{x}(t_n)\}$, find $\boldsymbol{f}$
Type	Forward problem	Inverse problem
Theory	Dalhquist etc	Keller-Du
Consistency	$\rho(1) = 0,\ \rho'(1) = \sigma(1)$	$\rho(1) = 0,\ \rho'(1) = \sigma(1)$
Stability	Dalhquist root condition on ρ	New root conditions on σ or $\hat{\sigma}$
Order	$\rho(e^z) - z\sigma(e^z) = O(z^{k+1})$	$\rho(e^z) - z\sigma(e^z) = O(z^{k+1})$
Examples	BDF-M (M$\leq$ 6), Adams-Family	BDF, AB-M (M$\leq$6), AM-M (M$\leq$2)

Note: the theory on learning dynamics is also valid, e.g. with function approximations (Du-Gu-Yang-Zhou 2021) such as ReLU FNNs, but the ill-posed nature of inverse problems remains unaccounted for.

Dynamics Discovery Using LMM

Key: zero LMM residual for idealized LMM based dynamics discovery, that is,

$$\sum_{m=m_0}^{M_0} \beta_m \hat{\boldsymbol{f}}_{n-m} = \frac{1}{h} \sum_{m=0}^{M} \alpha_m \hat{\boldsymbol{x}}_{n-m}, \quad M \le n \le N. \Leftrightarrow \text{Matrix form: } B\hat{\boldsymbol{f}}^h = h^{-1} A \hat{\boldsymbol{x}}^h$$

where m_0 and M_0 are defined by: $\beta_{m_0} \neq 0$ and $\beta_{M_0} \neq 0$ but $\beta_m = 0$, $\forall m \notin [m_0, M_0]$, and matrices A and B, respectively, encode $\{\alpha_m\}$ and $\{\beta_m\}$ (or $\{\beta_m\}_{m_0}^{M_0}$, to be precise).

Dynamics Discovery Using LMM

Key: zero LMM residual for idealized LMM based dynamics discovery, that is,

$$\sum_{m=m_0}^{M_0} \beta_m \hat{\boldsymbol{f}}_{n-m} = \frac{1}{h}\sum_{m=0}^{M} \alpha_m \hat{\boldsymbol{x}}_{n-m}, \quad M \le n \le N. \Leftrightarrow \text{Matrix form: } B\hat{\boldsymbol{f}}^h = h^{-1}A\hat{\boldsymbol{x}}^h$$

where m_0 and M_0 are defined by: $\beta_{m_0} \neq 0$ and $\beta_{M_0} \neq 0$ but $\beta_m = 0$, $\forall m \notin [m_0, M_0]$, and matrices A and B, respectively, encode $\{\alpha_m\}$ and $\{\beta_m\}$ (or $\{\beta_m\}_{m_0}^{M_0}$, to be precise).

Truncation error for the exact state and dynamics $(\boldsymbol{x}^h, \boldsymbol{f}^h)$: $\boldsymbol{\tau}^h = h^{-1}A\boldsymbol{x}^h - B\boldsymbol{f}^h$, i.e.,

$$(\boldsymbol{\tau}^h)_n = \frac{1}{h}\sum_{m=0}^{M} \alpha_m x(t_{n-m}) - \sum_{m=0}^{M} \beta_m f(x(t_{n-m})), \quad M \le n \le N. \qquad \text{(classical definition)}$$

For smooth state $x(t)$: $(\boldsymbol{\tau}_h)_n = \sum_{m=0}^{\infty} (-1)^m C_m h^{m-1} x^m(t_n)$, where $C_m = \sum_{k=1}^{M} \frac{k^m \alpha_k}{m!} + m \sum_{k=0}^{M} \frac{k^{m-1}\beta_k}{m!}$.

Convergence for dynamics discovery

Learned dynamics	$\hat{\boldsymbol{f}}^h$:	$B\hat{\boldsymbol{f}}^h = h^{-1}A\hat{\boldsymbol{x}}^h,$	for $\hat{\boldsymbol{x}}^h = \boldsymbol{x}^h$.
Exact dynamics	$\boldsymbol{f}^h$:	$B\boldsymbol{f}^h = h^{-1}A\boldsymbol{x}^h - \boldsymbol{\tau}^h,$	
Error on dynamics	$\boldsymbol{e}_d^h = \hat{\boldsymbol{f}}^h - \boldsymbol{f}^h$:	$B\boldsymbol{e}_d^h = h^{-1}A(\hat{\boldsymbol{x}}^h - \boldsymbol{x}^h) + \boldsymbol{\tau}^h = \boldsymbol{\tau}^h.$	

Convergence for dynamics discovery

Learned dynamics $\hat{\boldsymbol{f}}^h$: $B\hat{\boldsymbol{f}}^h = h^{-1}A\hat{\boldsymbol{x}}^h$, for $\hat{\boldsymbol{x}}^h = \boldsymbol{x}^h$.

Exact dynamics $\boldsymbol{f}^h$: $B\boldsymbol{f}^h = h^{-1}A\boldsymbol{x}^h - \boldsymbol{\tau}^h$,

Error on dynamics $\boldsymbol{e}_d^h = \hat{\boldsymbol{f}}^h - \boldsymbol{f}^h$: $B\boldsymbol{e}_d^h = h^{-1}A(\hat{\boldsymbol{x}}^h - \boldsymbol{x}^h) + \boldsymbol{\tau}^h = \boldsymbol{\tau}^h$.

Intuitively, as $h \to 0$, $\underbrace{\boldsymbol{\tau}^h \to 0}_{\text{Consistency}}$ and $\underbrace{\text{"}B^{-1}\text{" bounded}}_{\text{Stability}} \Rightarrow \underbrace{\boldsymbol{e}_d^h \to 0}_{\text{Convergence}}$ (to be made rigorous).

Convergence for dynamics discovery

Learned dynamics	$\hat{\boldsymbol{f}}^h$:	$B\hat{\boldsymbol{f}}^h = h^{-1}A\hat{\boldsymbol{x}}^h,$	for $\hat{\boldsymbol{x}}^h = \boldsymbol{x}^h$.
Exact dynamics	$\boldsymbol{f}^h$:	$B\boldsymbol{f}^h = h^{-1}A\boldsymbol{x}^h - \boldsymbol{\tau}^h,$	
Error on dynamics	$\boldsymbol{e}_d^h = \hat{\boldsymbol{f}}^h - \boldsymbol{f}^h$:	$B\boldsymbol{e}_d^h = h^{-1}A(\hat{\boldsymbol{x}}^h - \boldsymbol{x}^h) + \boldsymbol{\tau}^h = \boldsymbol{\tau}^h.$	

Intuitively, as $h \to 0$, $\underbrace{\boldsymbol{\tau}^h \to 0}_{\text{Consistency}}$ and $\underbrace{\text{"}B^{-1}\text{" bounded}}_{\text{Stability}} \Rightarrow \underbrace{\boldsymbol{e}_d^h \to 0}_{\text{Convergence}}$ (to be made rigorous).

In contrast, for time integration using LMM, let $\boldsymbol{e}_s^h = \boldsymbol{x} - \hat{\boldsymbol{x}}$, error for integrated state:

$$A\boldsymbol{e}_s^h = hB(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\hat{\boldsymbol{x}})) + h\boldsymbol{\tau}^h \approx h(B\nabla \boldsymbol{f})\boldsymbol{e}_s^h + h\boldsymbol{\tau}^h.$$

Numerical ODE: as $h \to 0$, $\underbrace{\boldsymbol{\tau}^h \to 0}_{\text{Consistency}}$ and $\underbrace{\text{"}A^{-1}\text{" bounded}}_{\text{Stability}} \Rightarrow \underbrace{\boldsymbol{e}_s^h \to 0}_{\text{Convergence}}$ (classical theory).

Consistency and stability for dynamics discovery

Definition (Consistency for dynamics discovery). As $h \to 0$:

consistent if $\|\boldsymbol{\tau}^h\|_\infty = \max_{M \le n \le N} |\boldsymbol{\tau}_n^h| \to 0$; strongly consistent if $\|\boldsymbol{\tau}^h\|_1 = \sum_{n=M}^{N} |\boldsymbol{\tau}_n^h| \to 0$.

Definition (Stability for Dynamics Discovery). As $h \to 0$, wrt prescribed IC in $I_M = \{i\}_{M-M_0}^{M-m_0-1}$:

stable if $\|\hat{\boldsymbol{f}}\|_\infty \lesssim \max_{i \in I_M} |\hat{\boldsymbol{f}}_i| + \|B\hat{\boldsymbol{f}}\|_\infty$; marginally stable if $\|\hat{\boldsymbol{f}}\|_\infty \lesssim \max_{i \in I_M} |\hat{\boldsymbol{f}}_i| + \|B\hat{\boldsymbol{f}}\|_1$.

Consistency and stability for dynamics discovery

Definition (Consistency for dynamics discovery). As $h \to 0$:

consistent if $\|\boldsymbol{\tau}^h\|_\infty = \max_{M \le n \le N} |\boldsymbol{\tau}_n^h| \to 0$; strongly consistent if $\|\boldsymbol{\tau}^h\|_1 = \sum_{n=M}^{N} |\boldsymbol{\tau}_n^h| \to 0$.

Definition (Stability for Dynamics Discovery). As $h \to 0$, wrt prescribed IC in $I_M = \{i\}_{M-M_0}^{M-m_0-1}$:

stable if $\|\hat{\boldsymbol{f}}\|_\infty \lesssim \max_{i \in I_M} |\hat{\boldsymbol{f}}_i| + \|B\hat{\boldsymbol{f}}\|_\infty$; marginally stable if $\|\hat{\boldsymbol{f}}\|_\infty \lesssim \max_{i \in I_M} |\hat{\boldsymbol{f}}_i| + \|B\hat{\boldsymbol{f}}\|_1$.

Definition (Root Conditions). $\sigma(r) = r^{M-M_0+m_0}\hat{\sigma}(r)$, $\{r_j\} = \{\text{roots of } \hat{\sigma}\} = \{\text{roots of } \sigma\} \backslash \{0\}$:

strong root condition: $|r_j| < 1, \forall j$; root condition: $|r_j| \le 1, \forall j$ and if $|r_j| = 1$, then it is simple.

Remark: root condition on $\rho(r)$ for time integration ($A\boldsymbol{e}_s \approx h\boldsymbol{\tau}^h$), and on $\sigma(r)$ for learning ($B\boldsymbol{e}_d^h = \boldsymbol{\tau}^h$).

Stability via root condition and convergence

Theorem (Stability for Discovery, Keller-Du 19)

stable $\Leftrightarrow$ *strong root condition for* σ *or* $\hat{\sigma}$; *marginally stable* $\Leftrightarrow$ *root condition for* σ *or* $\hat{\sigma}$.

Remark. There have been many studies in the literature on roots of ρ, but little on roots of σ (the latter plays no role in integrating known dynamics, nor in learning if $\hat{\boldsymbol{f}}^h$ is directly provided rather than solved for),

Stability via root condition and convergence

Theorem (Stability for Discovery, Keller-Du 19)

stable $\Leftrightarrow$ *strong root condition for* σ *or* $\hat{\sigma}$; *marginally stable* $\Leftrightarrow$ *root condition for* σ *or* $\hat{\sigma}$.

Remark. There have been many studies in the literature on roots of ρ, but little on roots of σ (the latter plays no role in integrating known dynamics, nor in learning if $\hat{\boldsymbol{f}}^h$ is directly provided rather than solved for),

Definition (Convergence for Discovery) $\|\boldsymbol{f}^h - \hat{\boldsymbol{f}}^h\|_\infty \to 0$ as $h \to 0$ and $\max_{i \in I_M} |\boldsymbol{f}_i - \hat{\boldsymbol{f}}_i| \to 0$.

Theorem (Convergence for Dynamics Discovery, Keller-Du 19)

consistency $+$ *stability* $\Rightarrow$ *convergence; strong consistency* $+$ *marginal stability* $\Rightarrow$ *convergence.*

Remark. A partial analog to the classical Dahlquist equivalence: convergence$\Rightarrow$consistency $+$ marginal stability.

Remark. regularizations may offer alternatives to the theory without auxiliary conditions $\max_{i \in I_M} |\boldsymbol{f}_i - \hat{\boldsymbol{f}}_i| \to 0$.

Other related discussions

Other alternative and refined notions for convergence (unpublished):

- Rationalized marginal stability: $|r_j| \leq 1$ and $\forall |r_j| = 1 \Rightarrow$ simple and $\exists K_j \in \mathbb{N} \Rightarrow r_j^{K_j} = 1,\ r_j \neq 1$.
- Enhanced consistency: $\max_{M \leq n \leq N} |\boldsymbol{\tau}_n^h| \to 0$ and $N \max_{M < n \leq N} |\boldsymbol{\tau}_n^h - \boldsymbol{\tau}_{n-1}^h| \to 0$.
- Enhanced consistency + Rationalized marginal stability $\Rightarrow$ Convergence.

Other related discussions

Other alternative and refined notions for convergence (unpublished):

- Rationalized marginal stability: $|r_j| \leq 1$ and $\forall |r_j| = 1 \Rightarrow$ simple and $\exists K_j \in \mathbb{N} \Rightarrow r_j^{K_j} = 1,\ r_j \neq 1$.
- Enhanced consistency: $\max_{M \leq n \leq N} |\boldsymbol{\tau}_n^h| \to 0$ and $N \max_{M < n \leq N} |\boldsymbol{\tau}_n^h - \boldsymbol{\tau}_{n-1}^h| \to 0$.
- Enhanced consistency + Rationalized marginal stability $\Rightarrow$ Convergence.

Order of convergence for discovery (subject to suitable approximations to initial data):

- Stable LMM: order of convergence = order of truncation error.
- Marginally stable LMM: order of convergence $\geq$ order of truncation error -1.

  but AM-1 has its order of convergence = 2, due to (rationalized) marginal stability.
- Optimal order (analog of Dahlquist 1st barrier): $M + 2$? ρ and σ both have only unit roots.

LMM: integrating and learning dynamics

Mathematical theory on LMM for integrating and learning dynamics		
Task	Integrating dynamics	Learning dynamics
Goal	Given $\boldsymbol{f}$, find $\boldsymbol{x}(t)$	Given $\{\boldsymbol{x}(t_n)\}$, find $\boldsymbol{f}$
Type	Forward problem	Inverse problem
Theory	Dalhquist etc	Keller-Du
Consistency	$\rho(1) = 0, \rho'(1) = \sigma(1),$	$\rho(1) = 0, \rho'(1) = \sigma(1),$
Stability	Dalhquist root condition on ρ	New root condition on σ or $\hat{\sigma}$
Order	$\rho(e^z) - z\sigma(e^z) = O(z^{k+1})$	$\rho(e^z) - z\sigma(e^z) = O(z^{k+1})$
Examples	Euler, BDF-M (M$\leq$ 6), Adams-Family	? (check stability!)

Popular LMMs: BDF, Adams-Moulton (AM), Adams-Bashforth (AB)

BDF $\hat{\sigma}(r) = 1$ (no root, stable) $\Rightarrow$ Theorem. BDFs are convergent.

Popular LMMs: BDF, Adams-Moulton (AM), Adams-Bashforth (AB)

BDF $\hat{\sigma}(r) = 1$ (no root, stable) $\Rightarrow$ Theorem. BDFs are convergent.

AB-1 $\hat{\sigma}(r) = 1$ (no root, stable)

AB-2 $\hat{\sigma}(r) = \frac{3}{2}r - \frac{1}{2}$ $(r = \frac{1}{3}$, stable)

AB-3 $\hat{\sigma}(r) = \frac{23}{12}r^2 - \frac{16}{12}r + \frac{5}{12}$ $(|r|^2 = \frac{5}{23}$, stable) $\Rightarrow$

AB-4 $\hat{\sigma}(r) = \frac{55}{24}r^3 - \frac{59}{24}r^2 + \frac{37}{24}r - \frac{9}{24}$ (stable)

AB-5 $\sigma(r) = \frac{1901}{720}r^4 - \frac{2774}{720}r^3 + \frac{2616}{720}r^2 - \frac{1274}{720}r + \frac{251}{720}$ (stable)

Popular LMMs: BDF, Adams-Moulton (AM), Adams-Bashforth (AB)

BDF $\hat{\sigma}(r) = 1$ (no root, stable) $\Rightarrow$ **Theorem**. BDFs are convergent.

AB-1 $\hat{\sigma}(r) = 1$ (no root, stable)

AB-2 $\hat{\sigma}(r) = \frac{3}{2}r - \frac{1}{2}$ $(r = \frac{1}{3}$, stable)

AB-3 $\hat{\sigma}(r) = \frac{23}{12}r^2 - \frac{16}{12}r + \frac{5}{12}$ $(|r|^2 = \frac{5}{23}$, stable)

AB-4 $\hat{\sigma}(r) = \frac{55}{24}r^3 - \frac{59}{24}r^2 + \frac{37}{24}r - \frac{9}{24}$ (stable)

AB-5 $\sigma(r) = \frac{1901}{720}r^4 - \frac{2774}{720}r^3 + \frac{2616}{720}r^2 - \frac{1274}{720}r + \frac{251}{720}$ (stable)

$\Rightarrow$ AB-M: convergent, $M \leq 5$;

Convergent for all M?

Popular LMMs: BDF, Adams-Moulton (AM), Adams-Bashforth (AB)

BDF $\hat{\sigma}(r) = 1$ (no root, stable) $\Rightarrow$ Theorem. BDFs are convergent.

AB-1 $\hat{\sigma}(r) = 1$ (no root, stable)

AB-2 $\hat{\sigma}(r) = \frac{3}{2}r - \frac{1}{2}$ ($r = \frac{1}{3}$, stable)

AB-3 $\hat{\sigma}(r) = \frac{23}{12}r^2 - \frac{16}{12}r + \frac{5}{12}$ ($|r|^2 = \frac{5}{23}$, stable)

AB-4 $\hat{\sigma}(r) = \frac{55}{24}r^3 - \frac{59}{24}r^2 + \frac{37}{24}r - \frac{9}{24}$ (stable)

AB-5 $\sigma(r) = \frac{1901}{720}r^4 - \frac{2774}{720}r^3 + \frac{2616}{720}r^2 - \frac{1274}{720}r + \frac{251}{720}$ (stable)

$\Rightarrow$ AB-M: convergent, $M \leq 5$; Convergent for all M?

AM-0 $\sigma(r) = 1$ (no root, stable)

AM-1 $\sigma(r) = \frac{1}{2}r + \frac{1}{2}$ ($r = -1$, marginally stable)

AM-2 $\sigma(r) = \frac{5}{12}r^2 + \frac{8}{12}r - \frac{1}{12}$ ($r = \frac{-4 \pm \sqrt{21}}{5}$, unstable)

AM-3 $\sigma(r) = \frac{9}{24}r^3 + \frac{19}{24}r^2 - \frac{5}{24}r + \frac{1}{24}$ (unstable)

AM-4 $\sigma(r) = \frac{251}{720}r^4 + \frac{646}{720}r^3 - \frac{264}{720}r^2 + \frac{106}{720}r - \frac{19}{720}$ (unstable)

$\Rightarrow$

Popular LMMs: BDF, Adams-Moulton (AM), Adams-Bashforth (AB)

BDF $\hat{\sigma}(r) = 1$ (no root, stable) $\Rightarrow$ Theorem. BDFs are convergent.

AB-1 $\hat{\sigma}(r) = 1$ (no root, stable)

AB-2 $\hat{\sigma}(r) = \frac{3}{2}r - \frac{1}{2}$ $(r = \frac{1}{3}$, stable)

AB-3 $\hat{\sigma}(r) = \frac{23}{12}r^2 - \frac{16}{12}r + \frac{5}{12}$ $(|r|^2 = \frac{5}{23}$, stable)

AB-4 $\hat{\sigma}(r) = \frac{55}{24}r^3 - \frac{59}{24}r^2 + \frac{37}{24}r - \frac{9}{24}$ (stable)

AB-5 $\sigma(r) = \frac{1901}{720}r^4 - \frac{2774}{720}r^3 + \frac{2616}{720}r^2 - \frac{1274}{720}r + \frac{251}{720}$ (stable)

$\Rightarrow$ AB-M: convergent, $M \leq 5$; Convergent for all M?

AM-0 $\sigma(r) = 1$ (no root, stable)

AM-1 $\sigma(r) = \frac{1}{2}r + \frac{1}{2}$ $(r = -1$, marginally stable)

AM-2 $\sigma(r) = \frac{5}{12}r^2 + \frac{8}{12}r - \frac{1}{12}$ $(r = \frac{-4 \pm \sqrt{21}}{5}$, unstable)

AM-3 $\sigma(r) = \frac{9}{24}r^3 + \frac{19}{24}r^2 - \frac{5}{24}r + \frac{1}{24}$ (unstable)

AM-4 $\sigma(r) = \frac{251}{720}r^4 + \frac{646}{720}r^3 - \frac{264}{720}r^2 + \frac{106}{720}r - \frac{19}{720}$ (unstable)

$\Rightarrow$

AM-0 : convergent;

AM-1 (strongly consistent): convergent;

AM-M: unstable divergent for $M = 2, 3, 4$.

(this explains earlier experiments!)

For other M?

Loss of stability of AM-M for $M \geq 2$

A-M derived from Lagrange interpolation:

$$\beta_m = \frac{(-1)^m}{m!(M-m)!} \int_0^1 \prod_{\substack{i=0 \\ i \neq m}}^{M} (x + i - 1) \, dx \, .$$

Properties ($M \geq 2$):

- $\beta_0 > |\beta_M| > 0$.
- $(-1)^m \beta_m < 0$ for $m \geq 1$.

Loss of stability of AM-M for $M \geq 2$

A-M derived from Lagrange interpolation:

$$\beta_m = \frac{(-1)^m}{m!(M-m)!} \int_0^1 \prod_{\substack{i=0 \\ i \neq m}}^{M} (x+i-1)\; dx \,.$$

Properties ($M \geq 2$):

- $\beta_0 > |\beta_M| > 0$.
- $(-1)^m \beta_m < 0$ for $m \geq 1$.

Example: AM-4 $\sigma(r) = \frac{251}{720} r^4 + \frac{646}{720} r^3 - \frac{264}{720} r^2 + \frac{106}{720} r - \frac{19}{720}$

Loss of stability of AM-M for $M \geq 2$

A-M derived from Lagrange interpolation:

$$\beta_m = \frac{(-1)^m}{m!(M-m)!} \int_0^1 \prod_{\substack{i=0 \\ i \neq m}}^{M} (x+i-1)\, dx\,.$$

Properties ($M \geq 2$):

- $\beta_0 > |\beta_M| > 0$.
- $(-1)^m \beta_m < 0$ for $m \geq 1$.

Example: AM-4 $\sigma(r) = \frac{251}{720} r^4 + \frac{646}{720} r^3 - \frac{264}{720} r^2 + \frac{106}{720} r - \frac{19}{720}$

Key: $\beta_1 > \beta_0 \Rightarrow (-1)^M \sigma(-\infty) > 0$ and $(-1)^M \sigma(-\frac{\beta_1}{\beta_0}) = \sum_{m \geq 2} (-1)^{-m} \beta_m (\frac{\beta_1}{\beta_0})^m < 0$.

Recall the abnormal behavior for AM-M for $M \geq 2$: it might be rooted in this instability!

Loss of stability of AM-M for $M \geq 2$

Lemma

$\beta_1 > \beta_0 > 0$ for AM-M, $M \geq 2$ *(Note: $\beta_1 = \beta_0 = 1/2$ for AM-1).*

Proof. For $M = 2$, we have $\beta_1 = \frac{8}{12} > \beta_0 = \frac{5}{12} > 0$. In general,

$$\beta_1 > \beta_0 > 0 \Leftrightarrow \frac{M}{M+1} \int_0^1 \prod_{i=1}^{M-1} (x+i)dx > \int_0^1 \prod_{i=0}^{M-1} (x+i)dx > 0.$$

Assume true for some $M > 2$, then for $M+1$, we have

$$\frac{M+1}{M+2} \int_0^1 \prod_{i=1}^{M} (x+i)dx = \frac{M+1}{M+2} \int_0^1 \left(x \prod_{i=1}^{M-1} (x+i) + M \prod_{i=1}^{M-1} (x+i)dx \right)$$

$$> \frac{M+1}{M+2} \left(\int_0^1 \prod_{i=0}^{M-1} (x+i) + \frac{M(M+1)}{M} \prod_{i=0}^{M-1} (x+i)dx \right) = \frac{(M+1)(M+2)}{(M+2)} \int_0^1 \prod_{i=0}^{M-1} (x+i)dx$$

$$> (M+1) \int_0^1 \frac{x+M}{M+1} \prod_{i=0}^{M-1} (x+i)dx = \int_0^1 \prod_{i=0}^{M} (x+i)dx, \qquad \square$$

Roots of AB-M for $M \geq 1$

Properties of AB-M for $M \geq 2$,

- $\beta_0 = 0$, $\beta_1 > |\beta_M| > 0$ and $(-1)^m \beta_m < 0$ for $m \geq 1$.
- $\hat{\sigma}$ associated with AB-M always has a root inside the unit disc if $M \geq 2$.

($\Rightarrow$ For $M \geq 2$, AB-M is unstable for dynamics discovery wrt *terminal data*. Same for AM-M.)

Computationally verified: $\hat{\sigma}$ of AB-M has a root r with $|r| > 1$ for $7 \leq M \leq 20$.

Conjecture: all roots of $\sigma = \sigma(r)$ are inside the unit disk only if $M < 7$. Note that $\sigma(r) = \int_0^1 p(r, x)dx$

where $p = p(r, x)$ is the degree M polynomial interpolant in $x \in \mathbb{R}$ of r^{M+x} at $\{x_j = -j\}_{j=0}^{M}$,

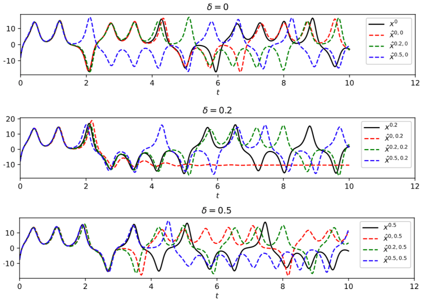
Extensions and open questions

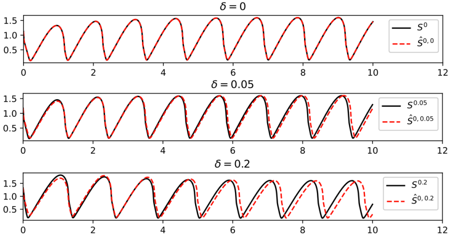
- Noisy data via smoothing of observed state with MLS (Keller-Du, unpublished);

Extensions and open questions

- Noisy data via smoothing of observed state with MLS (Keller-Du, unpublished);
- Convergence with function approximations (for learning/prediction), Du-Gu-Yang-Zhou, 2021);

  For stable LMM with aux conditions (initial, e.g.), the error is bounded by LMM truncation error and function approximation errors, for example, those with ReLU FNN. For unstable ones, DNN may still offer "reasonable" solutions, perhaps due to its ability to implicitly suppress oscillations.





DNNs also offer limited predictive capability:

Learning with simulated system with a given IC, predicting for different ICs via learned system.

Extensions and open questions

- Noisy data via smoothing of observed state with MLS (Keller-Du, unpublished);
- Convergence with function approximations (for learning/prediction), Du-Gu-Yang-Zhou, 2021);

Still many open (theoretical) questions

- Stable and optimal order LMMs for dynamics discovery (analog of Dahlquist 1st barrier)?
- Other schemes: Milne/Nystrom, predictor/corrector, multistage, structure-preserving integrators?
- General class(es) of LMMs that are convergent for both time integration and dynamics discovery?
- General linear/nonlinear multistep multistage integrators?
- Convergence with regularization such as minimum norm recovery, sparse learning?
- Inverse problem nature? non-uniqueness? implicit regularization? generalization error?

(could be suitable projects for students in 1st year numerical differential equation class)

Message of the day

⋆ Data-driven modeling of dynamic systems via machine learning is an inverse to integrating dynamics.

Model

Physics

Scientific Computation ($\mathrm{I^2SC}$)

Data

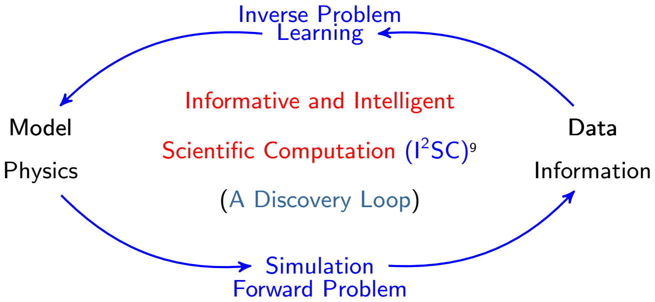
Information

Simulation

Forward Problem

Message of the day

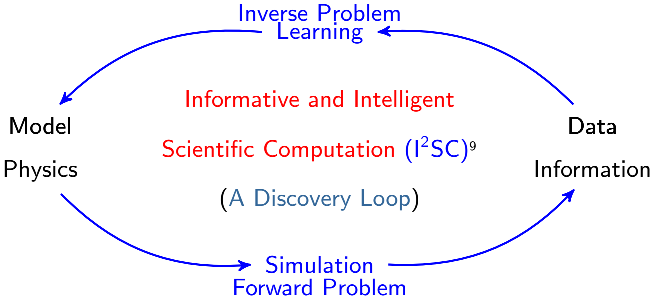
⋆ Data-driven modeling of dynamic systems via machine learning is an inverse to integrating dynamics.



[9] I^2SC was coined in 2004, published later in Du 2008, AIP Studies in Applied Math

Message of the day

⋆ Data-driven modeling of dynamic systems via machine learning is an inverse to integrating dynamics.



⋆ Avoid blind machine learning: what works for integrating dynamics might not be so for learning.

Thank you.

[9] I^2SC was coined in 2004, published later in Du 2008, AIP Studies in Applied Math