A Mass Conserving Machine Learning Algorithm for Atmospheric Chemistry

Obin Sturm and Anthony Wexler

Air Quality Research Center

University of California, Davis

UCDAVIS
AIR QUALITY RESEARCH CENTER

Many Thanks

Eladio Knipping and Naresh Kumar, EPRI,
for encouraging us to pursue this work

Some Prior Art

Potukuchi and Wexler, 1997

- Used NN to memorize time consuming part of aerosol thermodynamics
- 4-1200x speed improvement

Kelp, Tessum, Marshall, 2018

- Used NN to memorize $C(t+\Delta t)=f(C(t),T,RH,etc)$
- 250-4250x speed improvement
- Error propagation shortcomings

Keller & Evans, 2019

- Used Random Forest Regression
- 1.8x slower
- Error propagation shortcomings

Sources/Causes of Error

- Integrators don't conserve mass
  - Patches implemented to restore
- Thought experiment with NN integrator
  - Assume the integrated answer is "correct"
  - 0.1% error over each 6 minute time step
  - 168% error after a week (if error systematic)
- How much of the error propagation is due to lack of mass conservation?
- Our Goal: Derive a mass conserving framework for use with Machine Learning
  - Sorry for the math!

UCDAVIS
AIR QUALITY RESEARCH CENTER

The Math

$$\frac{\partial C}{\partial t} = F(C, T, RH, \text{actinic flux, stability, etc.})$$

$$F = AR$$

A = Stoichiometry matrix of the reactions

R = Reaction rates, kC_iC_j this and kC_iC_j that

Integrating gives

$$\Delta C_i = \sum_j A_{i,j} \int_t^{t+\Delta t} R_j(t)dt = \sum_j A_{i,j} S_j$$

Or in matrix form

$$\Delta C = AS$$

The Math

We need to solve

$$\Delta C = AS$$

So instead of memorizing

$$C(t + \Delta t) = C(t, T, RH, etc.)$$

We memorize

$$S(t \rightarrow t + \Delta t) = S(C(t), T, RH, etc.)$$

Advantages

- Mass conserving to machine precision
- S is simpler to memorize than C because fewer direct influences
- And because the complexity in A is removed

The Math Problem

How do we get data to memorize?

$$S(C(t), T, RH, etc.)$$

Via the inverse of

$$\Delta C = AS$$

Which is

$$S = A^{G,-1}\Delta C$$

where $A^{G,-1}$ is the generalized inverse of A

But A is rectangular so infinite number of generalized inverses

The solution:

Constrain $A^{G,-1}$ so that S lies in the "legal subspace"

$$A_S^{G,-1} = P_S(AP_S)^{G,-1}$$

P_S is a projection that defines the "legal subspace" for S

The Math Problem

P_S is defined as

$$P_S = U(U^+U)^{-1}U^+$$

where

$$U = \langle S_1|S_2|\ldots|S_i\rangle$$

The S_i are enough examples of "legal" S vectors to constrain $A^{G,-1}$

And we get them from

$$S_i \sim R_i(t)\Delta t$$

where the R are samplings of reaction rates from the mechanism

Conclusions

1. We have transformed the photochemistry problem into one that is mass conserving.

2. Current integrators are not necessarily mass conserving so potentially we have an advantage.

3. But the challenge still remains: Can machine learning algorithms perform well enough to replace integrators? Need much better than 0.1% accuracy.

4. We're on it.

5. Questions?