
ANITI
ARTIFICIAL & NATURAL INTELLIGENCE
TOULOUSE INSTITUTE
Data Assimilation with
Machine Learning
Certified
AI
Université
Fédérale
Toulouse
Midi-Pyrénées


ANITI
ARTIFICIAL & NATURAL INTELLIGENCE
TOULOUSE INSTITUTE
Data Assimilation with
Machine Learning
phys-constrained ML
Certified
AI
Université Fédérale
Toulouse Midi-Pyrénées

Outline

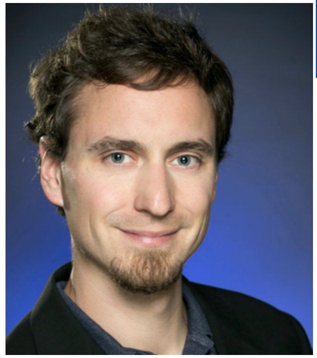
- **General presentation**
- **Some results**
  - Scientific results
  - Related works
  - Planned PhD / post doc proposals
- **Interaction with other chairs / industrial**

Chair members

ANITI
ARTIFICIAL & NATURAL INTELLIGENCE
TOULOUSE INSTITUTE

Serge Gratton

DA, optim.



Corentin Lapeyr

ML, physics

Selime Gurol

DA, optim.

Axel Carlier

ML, Images

Alfredo Buttari

LA, HPC

Pierre Boudier

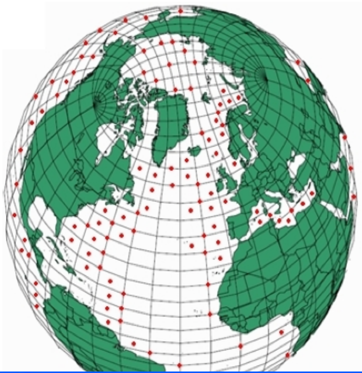
ML, GPUs

DA

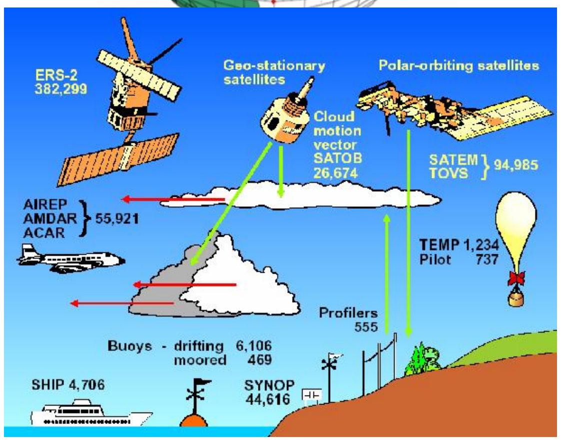
Model

$$
\begin{aligned}
\frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_z}{\partial z} &= 0 \\
\frac{\partial V_x}{\partial t} + V_x \frac{\partial V_x}{\partial x} + V_y \frac{\partial V_x}{\partial y} + V_z \frac{\partial V_x}{\partial z} &= -\frac{1}{\rho}\frac{\partial p}{\partial x} + \nu\left(\frac{\partial^2 V_x}{\partial x^2} + \frac{\partial^2 V_x}{\partial y^2} + \frac{\partial^2 V_x}{\partial z^2}\right) + f_x \\
\frac{\partial V_y}{\partial t} + V_x \frac{\partial V_y}{\partial x} + V_y \frac{\partial V_y}{\partial y} + V_z \frac{\partial V_y}{\partial z} &= -\frac{1}{\rho}\frac{\partial p}{\partial y} + \nu\left(\frac{\partial^2 V_y}{\partial x^2} + \frac{\partial^2 V_y}{\partial y^2} + \frac{\partial^2 V_y}{\partial z^2}\right) + f_y \\
\frac{\partial V_z}{\partial t} + V_x \frac{\partial V_z}{\partial x} + V_y \frac{\partial V_z}{\partial y} + V_z \frac{\partial V_z}{\partial z} &= -\frac{1}{\rho}\frac{\partial p}{\partial z} + \nu\left(\frac{\partial^2 V_z}{\partial x^2} + \frac{\partial^2 V_z}{\partial y^2} + \frac{\partial^2 V_z}{\partial z^2}\right) + f_z
\end{aligned}
$$

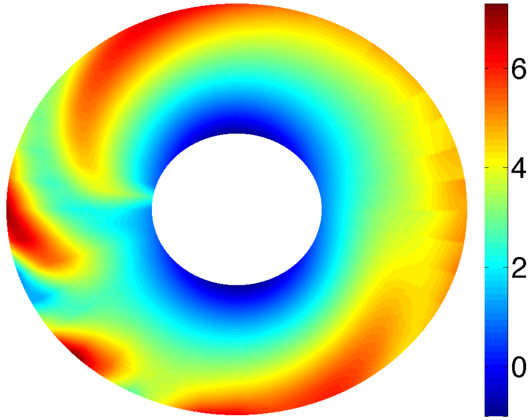
+



Data



Prediction



Implicit reduction and

What is Data Assimilation

ANITI
ARTIFICIAL & NATURAL INTELLIGENCE
TOULOUSE INSTITUTE

Historically, the purpose of data assimilation is to combine

- observations (measured, simulated,..)
- parameterized dynamical system model in order

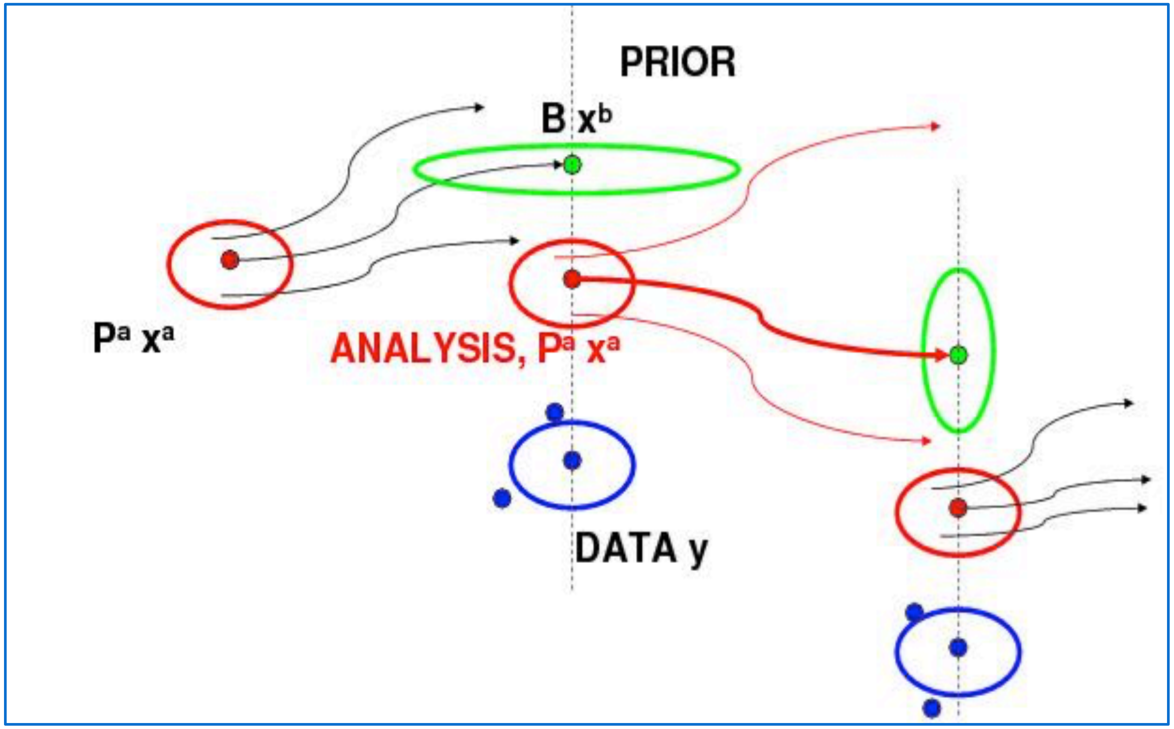
to produce better estimates of the current (future, sometimes "future in the past") state variables of the system → accuracy

In the variational data assimilation, information provided by the observations is used to find an optimal set of model parameters through a minimization process: there is a large scale optimization problem (10^6,sometimes even more variables) → HPC

There is a tendency to extend the assimilation procedure with Kalman filter and Monte Carlo techniques, to obtain useful sensitivity analysis or covariance analysis → applied statistics

Context

Large scale systems



$$y_0 = H(x_0) + \varepsilon_0$$

$$x_1 = M(x_0) + \varepsilon_1$$

$$\begin{cases} \text{Analysis} \quad p(x_0|y_0) \propto p\left(\varepsilon_0^r = y_0 - H(x_0)\right) \times p(x_0) \\ \text{Propagation} \quad p(x_1|y_0) \propto \int p\left(\varepsilon_1^q = x_1 - M(x_0)\right) \times p(x_0|y_0)\, dx_0 \end{cases}$$

Outline

- **General presentation**
- **Some results**
  - Scientific results
  - Related works
  - Planned PhD / post doc proposals
- **Interaction with other chairs / industrial**

Ensemble variational approach

ANITI
ARTIFICIAL & NATURAL INTELLIGENCE
TOULOUSE INSTITUTE

- Maximum likelihood approach in the Gaussian case leads to an optimization problem

$$\min_{x_0} \frac{1}{2}\left\|x_0 - x_b\right\|^2_{B^{-1}} + \frac{1}{2}\sum\left\|H\left(x\left(t_i\right)\right) - y_i\right\|^2_{R^{-1}} + \frac{1}{2}\sum\left\|x(t_{i+1}) - M_i\left(x\left(t_i\right)\right)\right\|^2_{Q^{-1}}$$

- In the Ensemble of DA approach, the observations and the background is sampled, leading to a set of optimization problems to solve.
- Problems/opportunities:
    - Concurrent applications of an optimization algorithm. ML and large scale optimization
    - Nonlinear models, their linearizations and adjoints are expensive to compute: ML for physical modelling
    - Additional DA tasks. ML for : covariance modelling, forecast sensivity to observation impact, bias detection
    - Embarquability issues

Optimisation

- Sequence of optimization problems

$$\min_x \frac{1}{2}\left\|x - x_b\right\|^2_{B^{-1}} + \frac{1}{2}\sum\left\|H\left(x(t_i)\right) - y_i\right\|^2_{R^{-1}} + \frac{1}{2}\sum\left\|x(t_{i+1}) - M_i\left(x(t_i)\right)\right\|^2_{Q^{-1}}$$

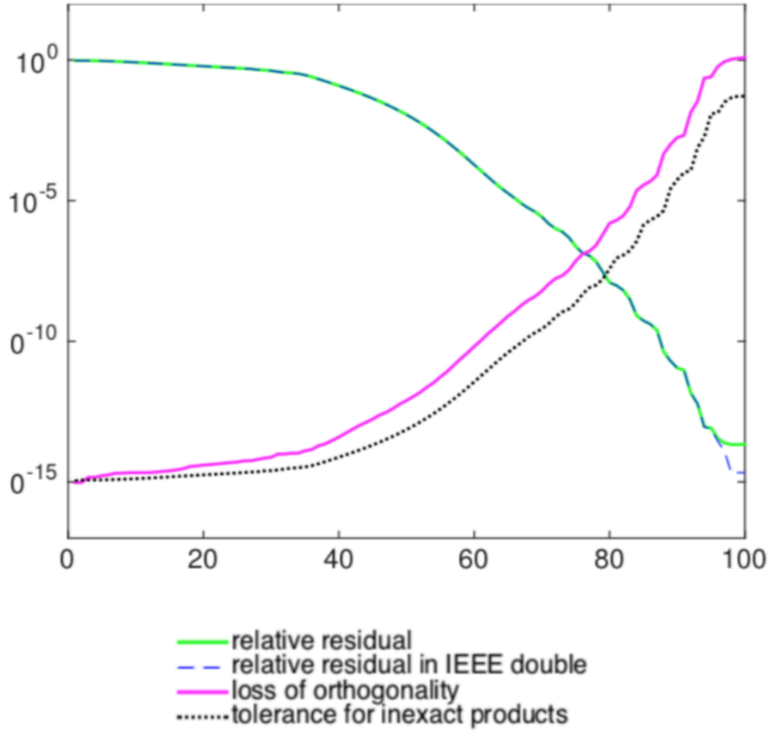
- Use of multiple precision arithmetics: solve he problem in single, half precision. Theory for handling multi-precision computations in optimization is being developed. Could also be used in stochastic gradient algorithms in GPUs?

- Already done:
    1. quadratic case: multi-precision conjugate gradient algorithm
    2. non-convex smooth case: theory + preliminary experiments
    3. non-convex + convex: composite case: in progress

Quadratic case (in progress)

Theoretical inexact CG algorithm

1. Set $x_0 = 0$, $\beta_0 = \|b\|_2^2$, $r_0 = -b$ and $p_0 = r_0$
2. For k=0, 1, . . . , do
3. $\quad c_k = (A + E_k)p_k$
4. $\quad \alpha_k = \beta_k / p_k^T c_k$
5. $\quad x_{k+1} = x_k + \alpha_k p_k$
6. $\quad r_{k+1} = r_k + \alpha_k c_k$
7. $\quad$ if r_{k+1} is small enough then stop
8. $\quad \beta_{k+1} = r_{k+1}^T r_{k+1}$
9. $\quad p_{k+1} = -r_{k+1} + (\beta_{k+1}/\beta_k)p_k$
10. EndFor



- Convergence is proved
- Important savings can be obtained in some applications
- Experiments need be done in DA

Smooth non-convex case

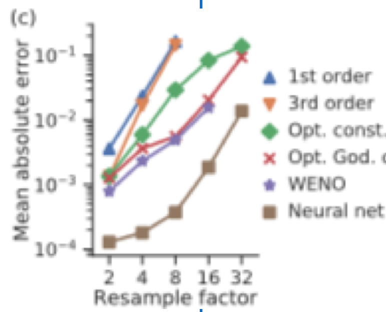
ANITI
ARTIFICIAL & NATURAL INTELLIGENCE
TOULOUSE INSTITUTE

						relative to LMQN		
ϵ	Variant	nsucc	its.	costf	costg	its.	costf	costg
1e-03	LMQN	82	41.05	42.04	42.04			
	iLMQN-a	80	50.05	9.88	6.11	1.23	0.24	0.15
	iLMQN-b	76	52.67	13.85	3.34	1.36	0.35	0.08

- Error in the function and in the gradient
- Trust region setting
- Main idea: the error in the function should be smaller than the model decrease
- Experiments need be done in DA too

Physical modelling

- ML can be useful for physical modelling and solving PDEs. How does this schemes work in DA?
- Data driven coarse graining of PDEs:
  - Use a super resolution neural net to predict physical quantities (such flux) at low resolution.
  - The net estimates are coefficients in WENO scheme
  - Only demonstrated for 1D + time PDE (Burgers..). Still expensive
  - Testing set forcing similar to training
- Assimilation in a latent space:
  - A network is used to link physical space to a latent space
  - The latent space variable are used for time marching the PDE
  - Write DA in such a system



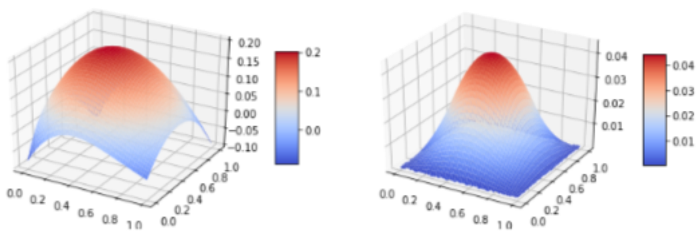
Optimization and PDE solving

- Several formulations are possible for solving a PDE with NN:
  - residual, variational formulations
  - coefficient to solution map
  - linear solution
- It is expected that NN dedicated architectures (tensor processors, or..) may increase the viability of the approach

- Standard solvers in NN framework (e.g., Tensorflow) can be to slow for some formulation
  - Second order methods can be faster (but expensive)
  - Multilevel Second order methods may be the way to go

$$-\Delta u - \left(\frac{2\pi\nu}{40}\right)^2 = g(x,y),$$

	iter	RMSE
ADAM	4000	1.e-1
MLM	200	1.e-3



ADAM MLM

Special topic in DA

- DA with machine learning. Do NN have a good shadowing property? Can NN (which architecture?) well represent the unstable modes, and have a good prediction property?

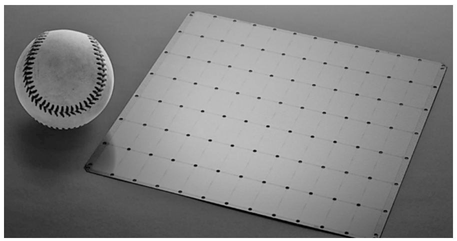
- The observation operator in DA can be computationally expensive.
  - Could DA be used to model this operator and its derivative. In simple cases it is just an interpolation operator: what is the quality of the NN for computing an optimization step

- FSOI. Some tasks in DA can be close to image processing, e.g., forecast sensitivity of observation impact estimation: reduces the cost of DA. Can this be done with NN?

Performance issues

Numerous performance issue arise due to the volume of data, complexity of algorithms and variety and complexity of computing platforms.

- **Large scale**
  - Performance and scalability on parallel systems
  - Efficient use of high performance computing platforms equipped with Multicores, GPUs, IA dedicated units (Nvidia Tensor Cores, Google TPUs, Intel Nervana)
- **Small scale**
  - Embedded systems
  - Energy consumption
  - Performance on limited processing and memory resources



Waferscale Cerebra

DGX-2

Outline

- **General presentation**
- **Some results**
  - Scientific results
  - Related works
  - Planned PhD / post doc proposals
- **Interaction with other chairs / industrial**

PhD and Postdoc

- PhD of Victor Marchais : Special topics in DA
    - Simplified observation operator
    - Forecast sensitivity analysis for observations
- PD Anthony Fillion. Introducing ML in ensemble Kalman Filters. Theoretical aspects
- Philippe Toint visiting in November for 1 week on optimization
- Collaboration with IRIT/APO team

- PhD(s) : computation of reduced model for computational physics and application to Data Assimilation
    - Use ML to fasten CFD solvers

Interaction with other chairs

- Truly multidisciplinary topic: many collaborations possible

- For example :
  - With J.M Loubes on change of distributions, fairness
  - With F. Gamboa on fundamental aspects on DA
  - With J. Bolte, J.B. Lasserre on large scale optimization
  - With N. Dobigeon for algorithms in image processing
  - With Th. Serre on memory-optimized learning algorithms