La puissance de calcul et les techniques de prévision pour les Geosciences en grande dimension

Support from RTRA-STAE network

Serge Gratton,
with E. Simon, Ph.L. Toint, L.N. Vicente, Z. Zhang
S. Gurol, P. Laloyau, M. Pontaud, A. Weaver

December, 2016
Outline

1. Generalities

2. Some perspectives in meteorology: a general view

3. Enhancing parallel performance by decoupling in time

4. Using hierarchy of observations

5. Subspace reduction
The art of forecasting...

A (dynamical) system is characterized by state variables, e.g.
- velocity components
- pressure
- density
- temperature
- gravitational potential

Goal: predict or understand the system behaviour at a future time from
- dynamical integration model
- observational data

Applications: climate, meteorology, oceanography, neutronics, finance, ...
→ forecasting problems
The **dynamical integration model** predicts the state of the system given the state at an earlier time.

→ integrating may lead to very large **prediction errors** (inexact physics, discretization errors, approximated parameters)

**Observational data** are used to improve accuracy of the forecasts.

→ but the data are **inaccurate** (measurement noise, under-sampling)

→ In atmosphere/ocean science millions of observations and variables are processed every day: **large scale inverse problem**
The **dynamical integration model** predicts the state of the system given the state at an earlier time.

→ integrating may lead to very large **prediction errors** (inexact physics, discretization errors, approximated parameters)

**Observational data** are used to improve accuracy of the forecasts.

→ but the data are **inaccurate** (measurement noise, under-sampling

→ In atmosphere/ocean science millions of observations and variables are processed every day: **large scale inverse problem**
starting from a priori knowledge on the state, the forward model is run:

*expensive in computer time*, not always very parallel
starting from a priori knowledge on the state, the forward model is run: expensive in computer time, not always very parallel

data are processed: screened, agglomerated
starting from a priori knowledge on the state, the forward model is run: expensive in computer time, not always very parallel

data are processed: screened, agglomerated

adjustment with respect to observations: best value to be found by some form of optimization
starting from a priori knowledge on the state, the forward model is run: expensive in computer time, not always very parallel

data are are processed: screened, agglomerated

adjustment with of model with respect to observations: best value to be found by some form of optimization

predictions are then issued
One possibility is to approximately solve a large-scale non-linear weighted least-squares problem:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|x - x_b\|^2_{B^{-1}} + \frac{1}{2} \sum_{j=0}^{N} \|\mathcal{H}_j(\mathcal{M}_j(x)) - y_j\|^2_{R_j^{-1}}$$

where

- $x \equiv x(t_0)$ is the control variable in $\mathbb{R}^n$, $n \sim 10^6$.
- $\mathcal{M}_j$ are model operators: $x(t_j) = \mathcal{M}_j(x(t_0))$
- $\mathcal{H}_j$ are observation operators: $y_j \approx \mathcal{H}_j(x(t_j))$
- the observations $y_j$ and the background $x_b$ are noisy
- $B$ and $R_j$ are covariance matrices

- Model integration and minimization are crucial mathematical steps to optimize to achieve good performance
- Intricate combination of physics, applied mathematics and computer science
- Steered by application challenges
Improving the forecasting capabilities of the system: case of Meteorology

The main goal is to obtain fast, accurate algorithms that yields a good representation of reality.

Raises challenges:

- Physical modelling. Use finer grids, incorporate as many observations as possible and constraints.

- Statistical modelling. Try to learn from earlier forecast exercises. Estimate rather than impose errors on a priori knowledge and dynamical model.

- Algorithmic. Solve large-scale systems, large-scale optimization problems.

- Machine. Fast machines based on sustainable models in terms of cost, consumption, with balanced performance.
Some challenges on the model side

- At Meteo-France, essentially 2 models: global and limited area/regional. A ratio of 3-5 between resolution of both.

- Want to increase resolution of limited area to:
  - take into account topography, soil moisture for fog forecast
  - Increase resolution also needed for cevenol events

- Want to add new variables to temperature, pressure, etc. For instance dust concentration, aerosols. Want to add also chemistry with new applications expected in air quality. This would result in an improved physics

- All these improvements ask for ever larger computer resources.

- Huge socio-economical impact
Challenges on the optimization side (I) : weak-constrained variational approach

\[
\min_{x \in \mathbb{R}^n} \frac{1}{2} \|x_0 - x_b\|_B^{-1}^2 + \frac{1}{2} \sum_{j=0}^{N} \|\mathcal{H}_j(x_j) - y_j\|_{R_j^{-1}}^2 + \frac{1}{2} \sum_{j=1}^{N} \|x_j - M_j(x_{j-1}) - q_j\|_{Q_j^{-1}}^2
\]

- \( x = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_N \end{pmatrix} \in \mathbb{R}^n \) is the control variable (with \( x_j = x(t_j) \))

- Notation same as before but :
  - Matrices \( Q_j \) are the covariance matrices for model error. They should be estimated
  - Dramatically increase in the number of degrees of freedom
  - Solution computational cost does not scale linearly with problem size unless efficient acceleration techniques are found

We are facing a large-scale weighted nonlinear least-squares problem
Parallelism in time: saddle Point Approach

Let us consider weak-constraint 4D-Var as a constrained problem:

$$\min_{(\delta p, \delta w)} \frac{1}{2} \| \delta p - b \|_{D^{-1}}^2 + \frac{1}{2} \| \delta w - d \|_{R^{-1}}^2$$

subject to \( \delta p = L \delta x \) and \( \delta w = H \delta x \)

We can write the Lagrangian function for this problem as

$$\mathcal{L}(\delta w, \delta p, \lambda, \mu) = \frac{1}{2} \| \delta p - b \|_{D^{-1}}^2 + \frac{1}{2} \| \delta w - d \|_{R^{-1}}^2 + \lambda^T (\delta p - L \delta x) + \mu^T (\delta w - H \delta x)$$

The stationary point of \( \mathcal{L} \) satisfies the following equations:

$$D^{-1}(L \delta x - b) + \lambda = 0 \quad (1)$$
$$R^{-1}(H \delta x - d) + \mu = 0 \quad (2)$$
$$L^T \lambda + H^T \mu = 0 \quad (3)$$
Preconditioning Saddle Point Formulation of 4D-Var

\[ A = \begin{pmatrix} D & 0 & L \\ 0 & R & H \\ L^T & H^T & 0 \end{pmatrix} = \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \]

- B is the most computationally expensive block and calculations involving A are relatively cheap.

The inexact constraint preconditioner proposed by (Bergamaschi et. al. 2005) is promising for our application. The preconditioner can be chosen as:

\[ P = \begin{pmatrix} A & B^T \\ \tilde{B} & 0 \end{pmatrix} = \begin{pmatrix} D & 0 & \tilde{L} \\ 0 & R & 0 \\ \tilde{L}^T & 0 & 0 \end{pmatrix}, \]

where

- \( \tilde{L} \) is an approximation to the matrix L
- \( \tilde{B} = [\tilde{L}^T \ 0] \) is a full row rank approximation of the matrix \( B \in \mathbb{R}^{n \times (m+n)} \)
At each iteration the forcing formulation requires one application of $L^{-1}$, followed by one application of $L^{-T}$ (16 sequential subwindow integrations).
Multilevel algorithm in the dual space

Motivations

- "Huge" amount of data (even if the system is under sampled).
  - Assimilation computationally expensive.
- Heterogeneous spatial distribution of the observation.
  - Numerous observations in some areas VS few observations in some others.
- Do we need to assimilate all the observations to reach a target accuracy?

Fine and coarse subproblems

- The fine observation grid data assimilation problem:

\[
\min_{\delta x_f \in \mathbb{R}^n} \frac{1}{2} \| x + \delta x_f - x_b \|_{B^{-1}}^2 + \frac{1}{2} \| H_f \delta x_f - d_f \|_{R_f^{-1}}^2
\]

\[
(\delta x_f, \lambda_f) \text{ s.t. } \begin{cases} 
(R_f^{-1} H_f B H_f^T + I_{mf}) \lambda_f = R_f^{-1} (d_f - H_f (x_b - x)) \\
\delta x_f = x_b - x + BH_f^T \lambda_f 
\end{cases}
\]
An example of observation sets

Coarse observation set $\mathcal{O}_c$

Auxiliary observation set $\tilde{\mathcal{O}}_f$
Example: the Lorenz-96 system

Configuration of the experiment

- Model
  - $u$ is a vector of $N$-equally spaced entries around a circle of constant latitude.
  - Chaotic behavior for $F > 5$ and $N > 11$.

$$\forall j \in \mathbb{N}_N, \ \theta \in \mathbb{N}_\Theta, \ \frac{du_{j+\theta}}{dt} = \frac{1}{\kappa} \left( -u_{j+\theta-2} u_{j+\theta-1} + u_{j+\theta-1} u_{j+\theta+1} - u_{j+\theta} + F \right)$$

$u_N = u_0; \ u_{-1} = u_{N-1}; \ u_{N+1} = u_1$

with $N = 40$, $F = 8$, $\kappa = 120$ and $\Theta = 10$, $T = 120$ and $\Delta t = \frac{1}{80}$

- Background and observations
  - Normal distributed additive noise:
    - $\mathcal{N}(0, \sigma_{b/o}^2)$ with $\sigma_b = 0.2$, $\sigma_o = 0.1$.
  - $B = \sigma_b^2 l_n$ and $R = \sigma_o^2 l_p$. 
Example : Cost function and RMS error

La puissance de calcul et les techniques de prévision pour les Geosciences en grande dimension
Hybrid approach : Envar

- An ensemble method enables to compute a subspace that contains the direction that explain most of the variability of the system.
- The variational problem is solved on the linear space spanned by ensembles. Not very different from model reduction techniques : need subspace selection.
- No need to compute derivatives.
- *Embarrassingly* parallel!

Shallow water equations:

\[
\begin{align*}
\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} - fv + g \frac{\partial z}{\partial x} &= \nu \Delta u \\
\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + fu + g \frac{\partial z}{\partial y} &= \nu \Delta v \\
\frac{\partial z}{\partial t} + \frac{\partial uz}{\partial x} + \frac{\partial vz}{\partial y} &= \nu \Delta z
\end{align*}
\]
Uncertainty quantification

Ask more to the prediction system:

- Want to estimate errors in initial conditions, in models, in boundary conditions and their impact
- Associate a measure of fidelity/probability to the forecast

Run an ensemble of predictions. Several proposal by the community (Ensembles of 4D-Var or 4D-Envar):

- A set of model trajectories is computed
- A filter is run on a subspace span by the ensembles using the observation and model likelihoods, in a small dimensional space

Raises new opportunity in terms of error spread analysis
Ensemble/particule filters

Good news:

- are based on a sophisticated **sampling** of the probability densities involved in the inverse problem
- are naturally **embarrassingly** parallel.
- are based on a **derivative free** variants of the Kalman filter. Do not need a sophisticated tangent or adjoint code

Not so good news:

- suffer from the **curse of dimensionality**
- need techniques to alleviate the sampling error: **localization, inflation, colapse of weights**
- are based on a derivative free variant of the Kalman filter. Do not need a sophisticated tangent or adjoint code

Raise a number of important issues and stimulate new ideas: **critical sets, quasi-static approaches**
Conclusion/Remarks

- **New applications**, like uncertainty quantification call for a new generation of algorithms.
- Most applications are really large scale and use of parallel computers is mandatory. Issues like performance evaluation, fault tolerance, energy consumption should be considered at first place.
- Completely new (at least for the domain) algorithms have to be designed to face these challenges. Parallel in time, multigrid, saddle points, uncertainty quantification, global optimization.
- As usual, available computers also favour particular algorithms according to their hardware particularities.
- Exploiting the partially separable nature in ensemble methods may be useful.
- Hybridization of particule filters and variational approaches.

La puissance de calcul et les techniques de prévision pour les Geosciences en grande dimension