Observations Thinning in Data Assimilation

Philippe Toint
(with Serge Gratton, Monserrat Rincon-Camacho and Ehouarn Simon)

Department of Mathematics, University of Namur, Belgium

(philippe.toint@unamur.be)

San Diego, May 2014

Motivation: data assimilation for weather forecasting

(Attempt to) predict. . .

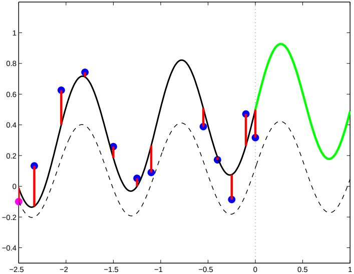
- tomorrow's weather
- the ocean's average temperature next month
- future gravity field
- future currents in the ionosphere
- . . .

using observations of continuous variables.

Data assimilation for weather forecasting (3)

The principle:

Minimize deviation between model and past observations



temp. vs. days

- Known situation 2.5 days ago and background prediction
- Record temperature for the past 2.5 days
- Run the model to minimize difference between model and observations
- Predict temperature for the next day

$$\min_{x_0} \frac{1}{2}\|x_0 - x_b\|^2_{B^{-1}} + \frac{1}{2}\sum_{i=0}^{N} \|\mathcal{H}\mathcal{M}(t_i, x_0) - b_i\|^2_{R_i^{-1}}.$$

Data assimilation problem: reformulation

The 4DVAR formulation for

$$\min_{x_0} \frac{1}{2}\|x_0 - x_b\|^2_{B^{-1}} + \frac{1}{2}\sum_{i=0}^{N} \|\mathcal{H}\mathcal{M}(t_i, x_0) - y_i\|^2_{R_i^{-1}}.$$

Gauss-Newton method : iterate on

1. linearize, concatenate successive times, define $x_0 = x + s$
2. solve

$$\boxed{\min_{x_0} J(s) \stackrel{\text{def}}{=} \tfrac{1}{2}\|x + s - x_b\|_{B^{-1}} + \tfrac{1}{2}\|Hs - d\|^2_{R^{-1}}}$$

where H is $m \times n$

Primal and dual approaches

Three possible approaches for solving the subproblem :

- the primal method : apply (preconditioned) truncated conjugate-gradients for s (dimension n)
- Use duality to re-write the problem as

$$\min_s J(s) \stackrel{\text{def}}{=} \tfrac{1}{2}\|x+s-x_b\|^2_{B^{-1}} + \tfrac{1}{2}\|a\|^2_{R^{-1}}$$

such that $$a = Hs - d$$

write the corresponding KKT conditions:

$$(R + HBH^T)\lambda = d - Hc, \quad s = c + BH^T\lambda, \quad a = -R\lambda$$

and apply (preconditioned) truncated conjugate-gradients in the H^TBH inner product to the 1rst equation (dimension m). Then compute s from the 2nd. (RPCG, Gratton-Tshimanga)

Potential advantages of the dual approach

- mathematically equivalent to the primal approach (same iterates), with /without preconditioning
- easily truncated without compromising convergence of the Gauss-Newton algorithm ($\neq$ PSAS)
- computationally attractive when $m \ll n$

$\Longrightarrow$

favorable when the number of observations is (relatively) small

Note : multipliers λ measure the impact of each observation individually!

Observation thinning

In many applications,

- too many observations in some parts of the domain!
- observations can be considered into a nested hierarchy $\{\mathcal{O}_i\}_{i=0}^r$ with

$$\mathcal{O}_i \subset \mathcal{O}_{i+1} \quad i = 0, \ldots r-1.$$

(coarse vs fine)

Can we exploit this for reducing computations?

The coarse and fine subproblems

The fine (sub)problem:

$$\min_s \tfrac{1}{2}\| x + s_f - x_b \|_{B^{-1}} + \tfrac{1}{2}\| H_f s - d_f \|^2_{R_f^{-1}}$$

The coarse (sub)problem:

$$\min_s \tfrac{1}{2}\| x + s_c - x_b \|_{B^{-1}} + \tfrac{1}{2}\| \Gamma_f (H_f s - d_f) \|^2_{R_c^{-1}}$$

where Γ_f is the restriction from fine to coarse observations. Moreover

- fine problem formulation $\Longrightarrow$ fine multiplier λ_f
- coarse problem formulation $\Longrightarrow$ coarse multiplier λ_c

A useful error bound

Question: what is the difference between fine and coarse multipliers ?

If $\Pi_c = \sigma\Gamma_f$ is the prolongation from the coarse observations to the fine ones, then

$$\| \lambda_f - \Pi_c \lambda_c \|_{R_f + H_f B H_f^T} \leq \| d_f - H_f s_c - R_f \Pi_c \lambda_c \|_{R_f^{-1}}$$

(proof somewhat technical...)

- Uses d_f but no comptuted quantity at the fine level
- Observation i useful if the i-th component of $\lambda_f - \Pi_c \lambda_c$ is large

How to exploit this?

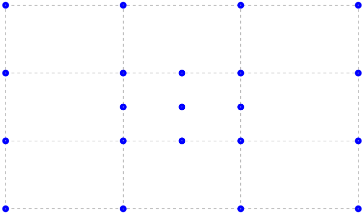
Idea:

Starting from the coarsest observation set, and until the finest observation set is used:

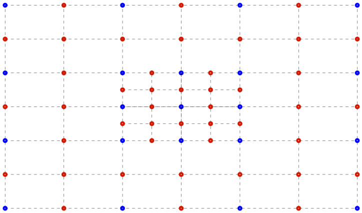
1. solve the coarse problem for (s_c, λ_c)
2. define a finer auxiliary problem by moving up in the hierarchy of observation sets (i.e. consider finer auxiliary observations)
3. use theorem to estimate distances from λ_c to $\lambda_{\rm aux} = \Pi_c \lambda_c$
4. using this, select a subset of the auxiliary observations whose impacts represents the impacts of these observations well enough (thinning)
5. redefine this selection as the next coarse observation set and loop

(needs: a more formal definition of the observations hierarchy
$+$ selection procedure)

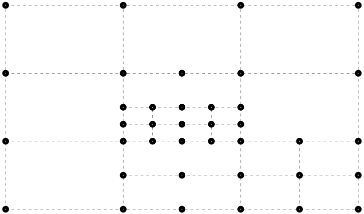
An example of observation sets



Coarse set



Auxilary set



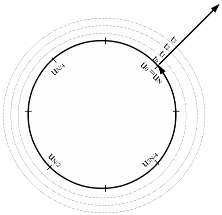
Selected fine set

Example 1: The Lorenz96 chaotic system (1)

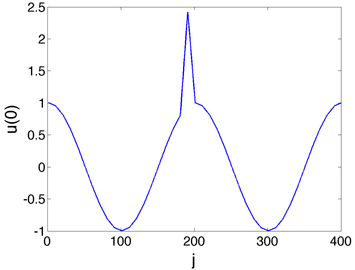
Find u_0, where $\bar{u}$ is an N-equally spaced entries around a circle, obeying

$$\frac{du_{j+\theta}}{dt} = \frac{1}{\kappa}(-u_{j+\theta-2}u_{j+\theta-1} + u_{j+\theta-1}u_{j+\theta+1} - u_{j+\theta} + F),$$

$(j = 1, \ldots 400, \;\; \theta = 1, \ldots, 120)$

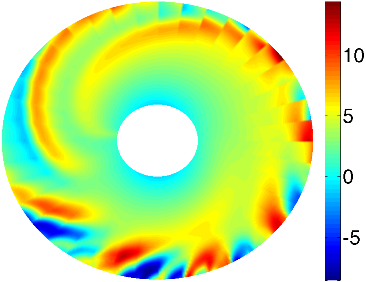


Coordinate system

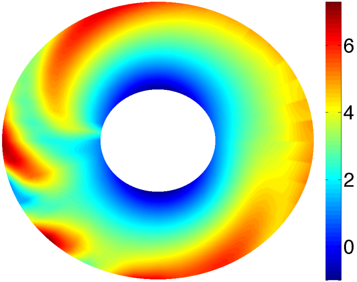


Initial $(u_1(0), u_2(0), \ldots, u_N(0))$

Example 1: The Lorenz96 chaotic system (2)

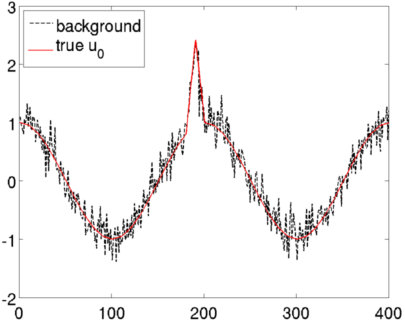


System over space and time

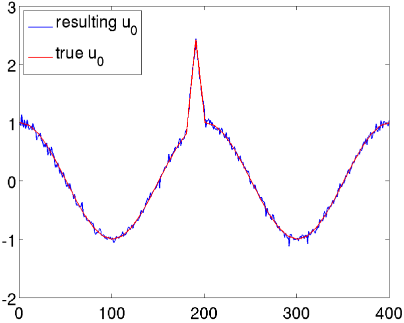


Window of assimilation

The Lorenz96 chaotic system (3)



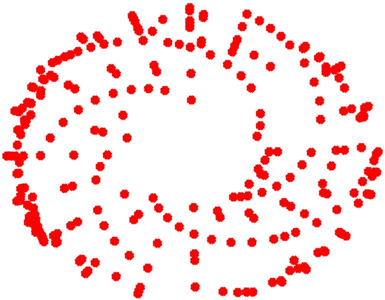
Background and true initial $u(0)$

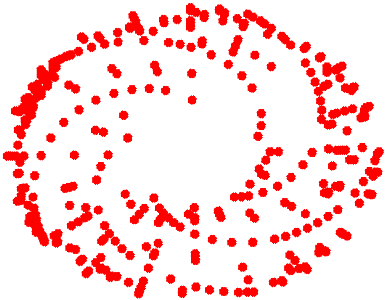


True and computed $u(0)$

Example 1: The Lorenz96 chaotic system (4)

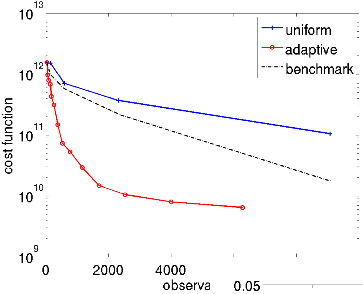
An example of transition from coarse to fine observations sets :



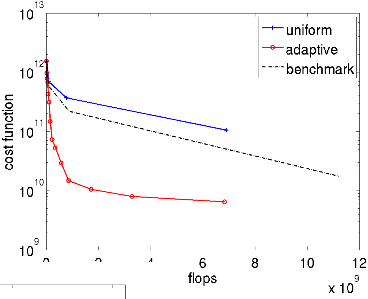


$\mathcal{O}_i \longrightarrow \mathcal{O}_{i+1}$

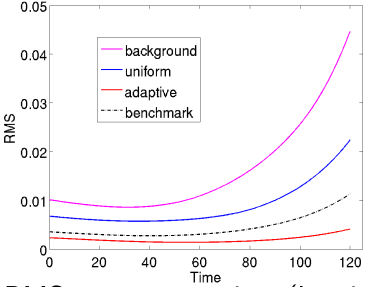
Example 1: The Lorenz96 chaotic system (5)



Cost vs obs



Cost vs flops



RMS error versus time (last iteration)

Conclusions and perspectives

Combining the advantages of the dual approach with adaptive observation thinning is possible

Observation thinning can produce faster solutions

Observation thinning can produce more accurate solutions

Reuse of selected data sets along the nonlinear optimization?

Use this idea for the design of observations campaigns?

Thank you for your attention!

Example 2: 1D wave system with a shock (1)

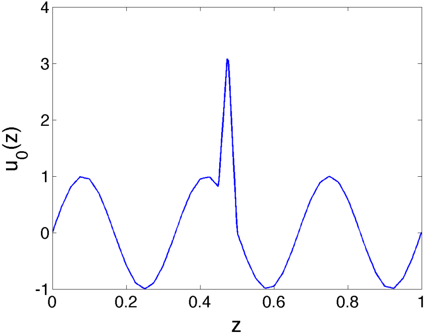
Find $u_0(z)$ in

$$\begin{aligned}
&\frac{\partial^2}{\partial t^2}u(z,t) - \frac{\partial^2}{\partial z^2}u(z,t) + f(u) = 0,\\
&u(0,t) = u(1,t) = 0,\\
&u(z,0) = u_0(z),\ \frac{\partial}{\partial t}u(z,0) = 0,\\
&0 \le t \le T,\ \ 0 \le z \le 1,
\end{aligned}$$

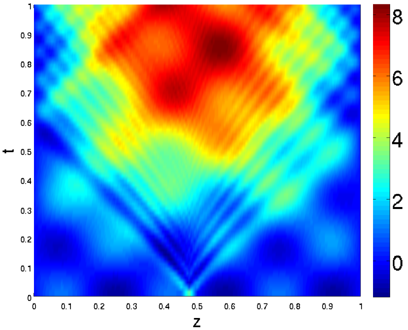
where $f(u) = \mu e^{\eta u}$

(360 grid points, $\Delta x \approx 2.8 \cdot 10^{-3}$, $T = 1$ and $\Delta t = \frac{1}{64}$).

Example 2: 1D wave system with a shock (2)

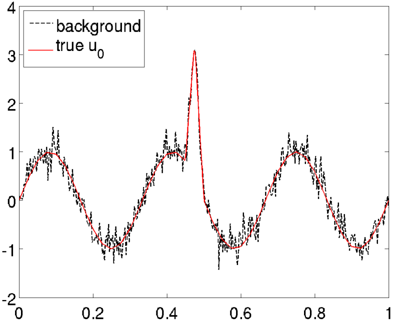


Initial $x_0 = u_0(z)$

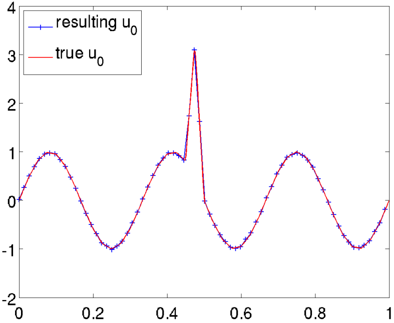


System over space and time

Example 2: 1D wave system with a shock (3)



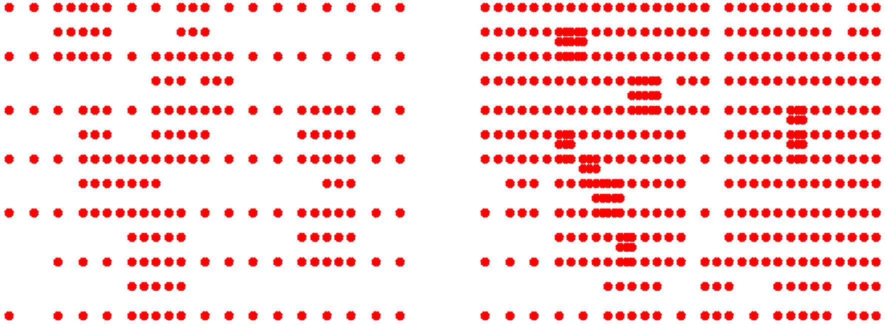
Background and true initial $u(0)$



True and computed $u(0)$

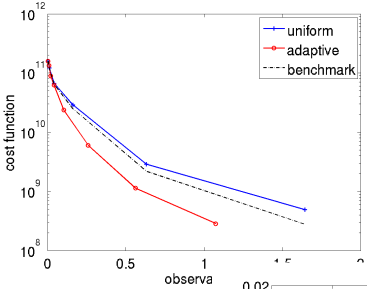
Example 2: 1D wave system with a shock (4)

An example of transition from coarse to fine observations sets :

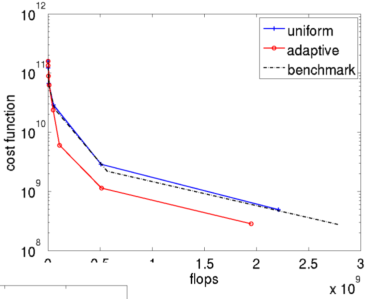


$\mathcal{O}_i \longrightarrow \mathcal{O}_{i+1}$

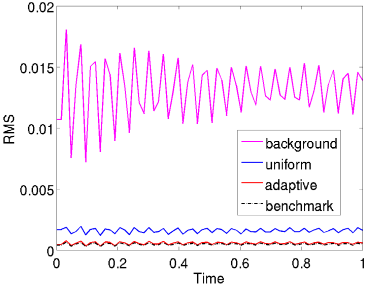
Example 1: 1D wave system with a shock (5)



Cost vs obs



Cost vs flops



RMS error versus time