Dual space multigrid strategies for variational data assimilation

Ehouarn Simon*
Serge Gratton, Monserrat Rincon-Camacho and Philippe Toint

* INPT, IRIT, Toulouse
ehouarn.simon@enseeiht.fr

1-5 June 2015

Variational data assimilation

Dual formulation

- Concatenation over time:

$$\min_{\delta x \in \mathbb{R}^n} \frac{1}{2}\|x - x_b + \delta x\|^2_{B^{-1}} + \frac{1}{2}\|H\,\delta x - d\|^2_{R^{-1}}$$

  with $H \in \mathcal{M}_{m,n}(\mathbb{R})$
- The problem can read:

$$\min_{\delta x \in \mathbb{R}^n} \frac{1}{2}\|x - x_b + \delta x\|^2_{B^{-1}} + \frac{1}{2}\|a\|^2_{R^{-1}}$$
$$\text{s.t. } a = H\,\delta x - d$$

- KKT conditions:
  - $(R^{-1}HBH^T + I_m)\lambda = R^{-1}(d - H(x_b - x)), \quad \delta x = x_b - x + BH^T\lambda$
- RPCG (Gratton and Tshimanga, 2009)
  - λ: apply (preconditioned) truncated conjugate gradient in the HBH^T inner product (dimension m).
  - Compute δx from λ.
  - Equivalent to the primal approach.
  - Easily truncated without compromising convergence of the GN algorithm.
- Computationally attractive when $m << n$.

Observation thinning

Motivations

- "Huge" amount of data (even if the system is under sampled).
  - ▷ Assimilation computationally expensive.
- Heterogenous 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?

Selection of observations

- Criteria
  - ▷ Do not assimilate the full data set.
  - ▷ Computationally tractable.
- Observations: a nested hierarchy $\{\mathcal{O}_i\}_{i=0}^{r}$ with

$$\forall i \in [0, r-1], \quad \mathcal{O}_i \subset \mathcal{O}_{i+1}$$

Outline

1. A "multigrid" observation thinning

2. Towards a multigrid dual solver?

Outline

A bigrid data assimilation problem (I)

Notations

- $\mathcal{O}_c$ the coarse observation set with m_c observations and $\mathcal{O}_f$ the fine observation set containing m_f observations such that $m_c < m_f$ and $\mathcal{O}_c \subset \mathcal{O}_f$.
- $\Gamma_f : \mathbb{R}^{m_f} \to \mathbb{R}^{m_c}$ a restriction operator from the fine observation space to the coarse one.
- Π_c the prolongation operator from the coarse observation space to the fine one such as $\Pi_c = \sigma_f \Gamma_f^T$ for some $\sigma_f > 0$.

Fine and coarse subproblems

- The fine observation grid data assimilation problem:

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

$\triangleright$ $(\delta x_f, \lambda_f)$ s.t. $\begin{cases} (R_f^{-1} H_f B H_f^T + I_{m_f})\lambda_f = R_f^{-1}(d_f - H_f(x_b - x)) \\ \delta x_f = x_b - x + B H_f^T \lambda_f \end{cases}$

A bigrid data assimilation problem (II)

Notations

- $\mathcal{O}_c$ the coarse observation set with m_c observations and $\mathcal{O}_f$ the fine observation set containing m_f observations such that $m_c < m_f$ and $\mathcal{O}_c \subset \mathcal{O}_f$.
- $\Gamma_f : \mathbb{R}^{m_f} \to \mathbb{R}^{m_c}$ a restriction operator from the fine observation space to the coarse one.
- Π_c the prolongation operator from the coarse observation space to the fine one such as $\Pi_c = \sigma_f \Gamma_f^T$ for some $\sigma_f > 0$.

Fine and coarse subproblems

- The coarse observation grid data assimilation problem:

$$\min_{\delta x_c \in \mathbb{R}^n} \frac{1}{2}\|x + \delta x_c - x_b\|^2_{B^{-1}} + \frac{1}{2}\|\Gamma_f (H_f \delta x_c - d_f)\|^2_{R_c^{-1}}$$

Or equivalently:

$$\min_{\delta x_c \in \mathbb{R}^n} \frac{1}{2}\|x + \delta x_c - x_b\|^2_{B^{-1}} + \frac{1}{2}\|\Pi_c^T (H_f \delta x_c - d_f)\|^2_{\bar{R}_c^{-1}}, \text{ with } \bar{R}_c^{-1} = (\frac{1}{\sigma_f})^2 R_c^{-1} \quad (2)$$

$\triangleright$ $(\delta x_c, \lambda_c)$ s.t. $\begin{cases} (\bar{R}_c^{-1} \Pi_c^T H_f B H_f^T \Pi_c + I_{m_c}) \lambda_c = \bar{R}_c^{-1} \Pi_c^T (d_f - H_f (x_b - x)) \\ \delta x_c = x_b - x + B H_f^T \Pi_c \lambda_c \end{cases}$

An a posteriori error bound

Theorem

Let δx_f be the solution to the fine problem and λ_f the corresponding Lagrange multiplier to the constraint. Analogously, let δx_c be the solution to modified coarse problem (2) and λ_c the corresponding Lagrange multiplier. Then the a posteriori error bound satisfies the inequalities

$$\begin{aligned} \|\lambda_f - \Pi_c \lambda_c\|^2_{R_f + H_f B H_f^T} &\leq \|d_f - H_f \delta x_c - R_f \Pi_c \lambda_c\|^2_{(R_f + H_f B H_f^T)^{-1}} \\ \|\lambda_f - \Pi_c \lambda_c\|^2_{R_f + H_f B H_f^T} &\leq \|d_f - H_f \delta x_c - R_f \Pi_c \lambda_c\|^2_{R_f^{-1}} \end{aligned}$$

Remarks

- $R_f + H_f B H_f^T$: difficult computation of the inverse in variational data assimilation ($B \sim$ complex matrix-vector operator).
- Bound: no need for the solution of the fine problem (λ_f or δx_f).
- Observations: "useful" if the associated components of $\lambda_f - \Pi_c \lambda_c$ are large.

How to construct $\mathcal{O}_f$ from $\mathcal{O}_c$?

Assumptions

- Coarse observation set: partition of the observation space in a finite number of cells $\{c_j\}_{j=1}^{p_c}$ of measures $\{w_j\}_{j=1}^{p_c}$.
- Auxiliary set $\tilde{\mathcal{O}}_f$: all observations in $\mathcal{O}_c$ with the addition of a single additional potential observation point in the interior of each cell.

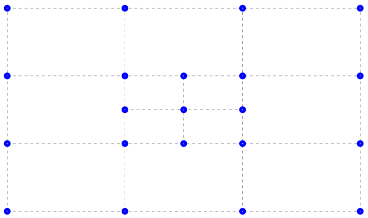
Selection

- Error indicator for each cell c_j of the auxiliary observation set $\tilde{\mathcal{O}}_f$

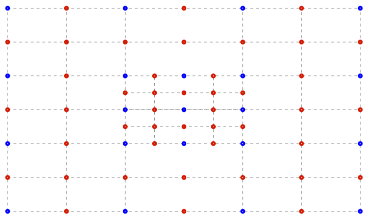
$$\forall j \in \mathbb{N}_p \quad \eta_j = w_j \left\langle (\tilde{d}_f - \tilde{H}_f \delta x_c - \tilde{R}_f \tilde{\Pi}_c \lambda_c)|_j, (\tilde{R}_f^{-1} (\tilde{d}_f - \tilde{H}_f \delta x_c - \tilde{R}_f \tilde{\Pi}_c \lambda_c))|_j \right\rangle$$

- Construction of a minimal set $\mathcal{S}_\eta$: $\theta \sum_{j=1}^{p} \eta_j \leq \sum_{k \in \mathcal{S}_\eta} \eta_k, \quad \theta \in (0,1)$
  - ▷ Priority to non-included cells with maximal error indicator values.
- $\mathcal{O}_f = \mathcal{O}_c \cup \left(\cup_{k \in \mathcal{S}_\eta} o_k \right)$

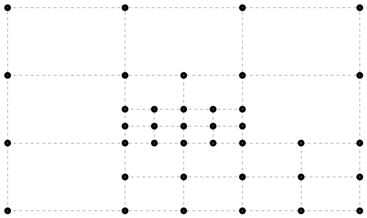
An example of observation sets



Coarse observation set $\mathcal{O}_c$



Auxiliary observation set $\tilde{\mathcal{O}}_f$



Fine observation set $\mathcal{O}_f$

Incremental 4D-Var with a multigrid observations thinning

1. Set $i = 0$, initialize x and the coarse observation set $\mathcal{O}_0$.
2. Find the solution $(\delta x_i, \lambda_i)$ to the problem
$$\min_{\delta x_i \in \mathbb{R}^n} \frac{1}{2}\|x_i + \delta x_i - x_b\|^2_{B^{-1}} + \frac{1}{2}\|H_i \delta x_i - d_i\|^2_{R_i^{-1}},$$
by approximately solving the system
$$(R_i^{-1} H_i B H_i^T + I_{m_i})\lambda_i = R_i^{-1}(d_i - H_i(x_b - x_i))$$
using RPCG and then setting $\delta x_i = x_b - x_i + B H_i^T \lambda_i$.
3. Given the set of observations $\mathcal{O}_i$, construct the auxiliary set $\tilde{\mathcal{O}}_{i+1}$.
4. For each cell c_j of observation set $\tilde{\mathcal{O}}_{i+1}$ compute the error indicators
$$\eta_j \quad = \quad w_j \, \langle (\tilde{d}_{i+1} - \tilde{H}_{i+1}\delta x_i - \tilde{R}_{i+1}\tilde{P}i_i\tilde{\lambda}_i)|_j, (\tilde{R}_{i+1}^{-1}(\tilde{d}_{i+1} - \tilde{H}_{i+1}\delta x_i - \tilde{R}_{i+1}\tilde{P}i_i\tilde{\lambda}_i))|_j \rangle$$
with $\tilde{\lambda}_i$ a modified Lagrange multiplier.
5. Build the set $\mathcal{S}_\eta$ such that
$$\theta_1 \left(\sum_{j=1}^{p_{i+1}} \eta_j \right) \leq \sum_{k \in \mathcal{S}_\eta} \eta_k$$
using the bulk chasing strategy.
6. Construct the set $\mathcal{O}_{i+1}$ as
$$\mathcal{O}_{i+1} := \mathcal{O}_i \cup \left(\bigcup_{k \in \mathcal{S}_\eta} o_k \right)$$
7. Update $x_i \leftarrow x_i + \delta x_i$, increment i and return to Step 2.

Example: the Lorenz-96 system

Configuration of the experiment

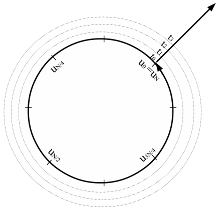
- Model
  - ▷ u is a vector of N-equally spaced entries around a circle of constant latitude.
  - ▷ Chaotic behavior for $F > 5$ and $N > 11$.

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

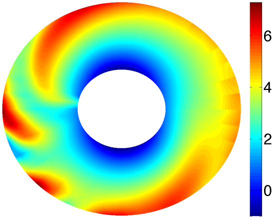
$$u_N = u_0; u_{-1} = u_{N-1}; u_{N+1} = u_1$$

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: $\mathcal{N}(0, \sigma^2_{b/o})$ with $\sigma_b = 0.2$, $\sigma_o = 0.1$.
  - ▷ $B = \sigma_b^2 I_n$ and $R = \sigma_o^2 I_p$.



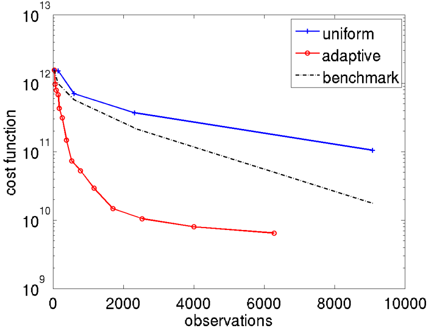
Coordinate system



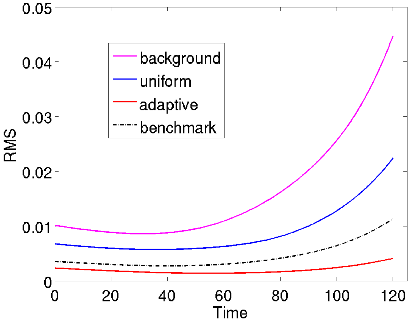
Dynamical system (space and time)

Example: Cost function and RMS error

Cost f. VS nb obs



RMS error VS time



Outline

1. A "multigrid" observation thinning
2. Towards a multigrid dual solver?

Multigrid methods for solving $Ax = b$ with iterative methods

Idea

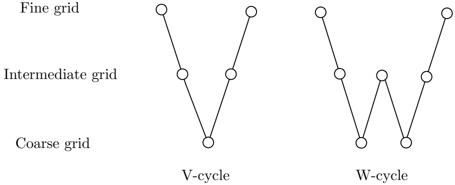
- Large scale components that are slow to converge on the high resolution grid may be reduced faster and at a smaller cost on a coarser resolution grid.
- Also applicable for nonlinear systems (Full Approximation Scheme; Brandt, 1982)

Two-level grids algorithm

- **Pre-smoothing**: Apply ν_1 steps of an iterative method S_1 on a fine grid
$$A_f x_f = b_f, \quad x_f = S_1^{\nu_1}(x_f, b_f)$$
- **Coarse grid correction**
  - Transfer the residual onto a coarser grid
  $$r_c = I_f^c(b_f - A_f x_f), \quad I_f^c : \text{ restriction operator}$$
  - Solve the problem on the coarse grid
  $$A_c \delta x_c = r_c$$
  - Transfer the correction onto the fine grid
  $$x_f = x_f + I_c^f \delta x_c, \quad I_c^f : \text{ interpolation operator}$$
- **Post-smoothing**: Apply ν_2 steps of an iterative method S_2 (most of the time identical to S_1) onto a fine grid
$$A_f x_f = b_f, \quad x_f = S_2^{\nu_2}(f x_f, b_f)$$

Multigrid methods for solving $Ax = b$ with iterative methods

Cycles



Convergence (Hackbusch, 2003)

- Smoothing property: smoothing steps should remove most of the error at small scales
  - Ellipticity of A (high frequencies associated to the largest eigenvalues).
  - Smoothing properties of the iterative solver.
- Prolongation/restriction operators: no amplification of the small scale components during a coarse correction step.
- Approximation properties: coarse grid correction steps should remove the error at large scales.
  - A_c close to A_f (discretization of the differential operator, $A_c = I_f^c A_f I_c^f$)

Multigrid methods in variational data assimilation

First-order necessary condition: $\nabla J(x) = 0$.

- Optimal control, constrained-PDE optimization: Brandt, Lewis and Nash (2005), Borzi and Schulz (2009).
- 4D-variational data assimilation: Neveu et al. (2011), Cioaca et al. (2013).
  - ▷ State space formulation.
- Dual space formulation: $A = HBH^T + R; \quad b = d - H(x_b - x)$.

Numerical experiments

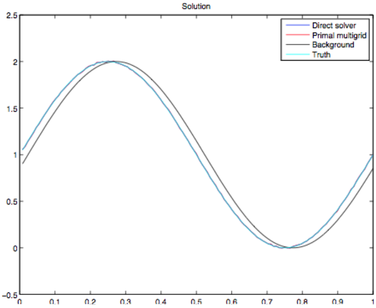
- Solution of a linear advection equation:

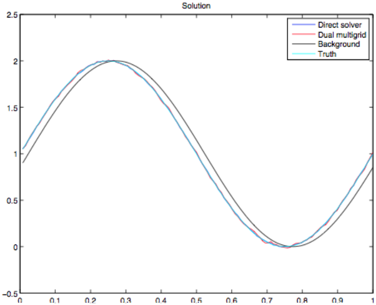
$$\frac{\partial u}{\partial t} + c\,\frac{\partial u}{\partial x} = 0, \text{ with } c > 0, x \in [0, L], t \in [0, T]$$

  with $c = 1\mathrm{m.s}^{-1}, L = 100\mathrm{m}, T = 78.125\mathrm{s}$

- Control variable: $u(t = 0)$.
- $B = \sigma_b^2 e^{-\frac{d^2}{Lcorr^2}}, \quad R = \sigma_o^2 I$.
- No observation thinning strategy: uniform observation grid at each level (3).

Numerical application

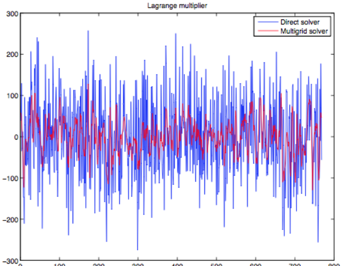




Primal approach: 1 V-cycle. Dual approach: 100 V-cycles.

Multigrid dual approach

- Increase of the residual after each coarse grid correction step.
- Conditions of convergence not fulfilled.

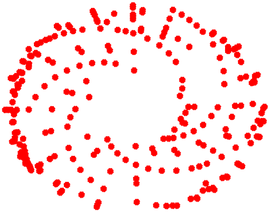


Conclusion and perspectives

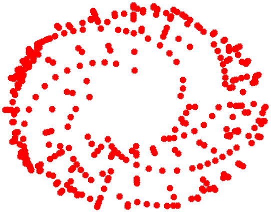
- A variational data assimilation approach combining observation thinning and dual-space conjugate-gradient techniques.
  - ▷ Exploiting the nested structure of the observations.
  - ▷ A posteriori error bounds based on Lagrange multipliers.
- Preliminary experiments.
  - ▷ Faster decrease of the cost function vs the amount of assimilated observations or flops.
- Preliminary experiments with a multigrid solver in dual space.
  - ▷ No improvement of the performances compared to an unigrid solver (even worst).
  - ▷ Characteristics of the problem not suitable for multigrid strategy? (Lagrange multipliers $\sim$ "noise")
- Further investigations
  - ▷ Modelling of the observation error covariance matrix properly taking into account the nested structure of the observations.

Thank you!

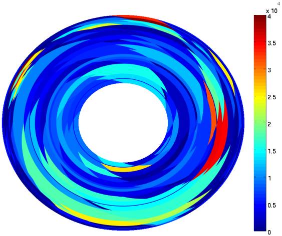
Example: Observation sets and adaptive errors



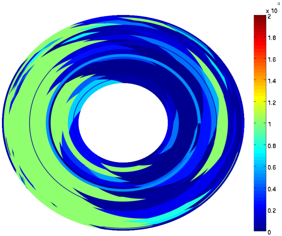
$\mathcal{O}_i$



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

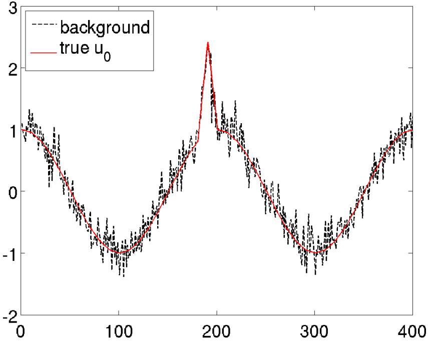


η_j

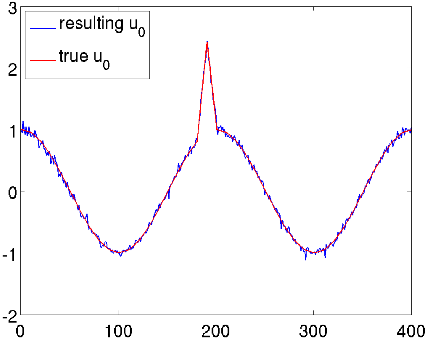


$$\epsilon_j = w_j \left\langle \Delta\lambda_{i+1}|_j, [(\tilde{R}_{i+1} + \tilde{H}_{i+1} B \tilde{H}_{i+1}^T)\Delta\lambda_{i+1}]|_j \right\rangle$$

Example: Control variable



Background vector and true $u(0)$



Algorithm solution and true $u(0)$

Example 2: 1D wave system with a shock

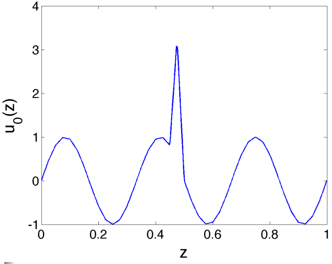
Configuration of the experiment

- Model

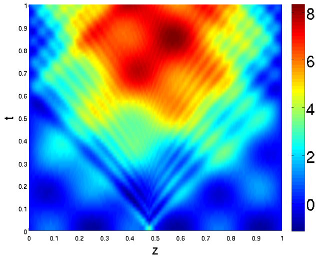
$$\frac{\partial^2}{\partial t^2}u(z,t) - \frac{\partial^2}{\partial z^2}u(z,t) + f(u) = 0,$$
$$u(0,t) = u(1,t) = 0,$$
$$u(z,0) = u_0(z),\ \frac{\partial}{\partial t}u(z,0) = 0,$$
$$0 \leq t \leq T,\ \ 0 \leq z \leq 1,$$

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

- Background and observations
  - ▷ Normal distributed additive noise: $\mathcal{N}(0, \sigma^2_{b/o})$ with $\sigma_b = 0.2$, $\sigma_o = 0.05$.
  - ▷ $B = \sigma_b^2 I_n$ and $R = \sigma_o^2 I_p$.

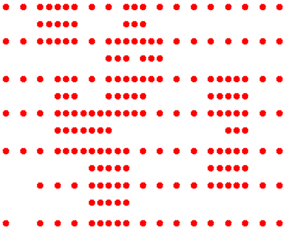


Initial $x_0 := u_0(z)$

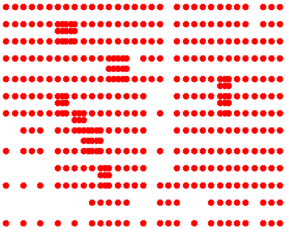


Dynamical system (space and time)

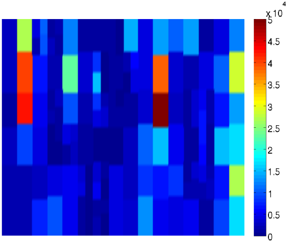
Example 2: Observation sets and adaptive errors



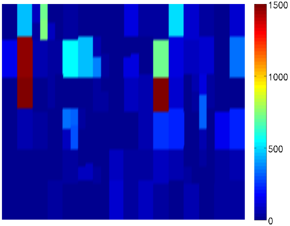
$\mathcal{O}_i$



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

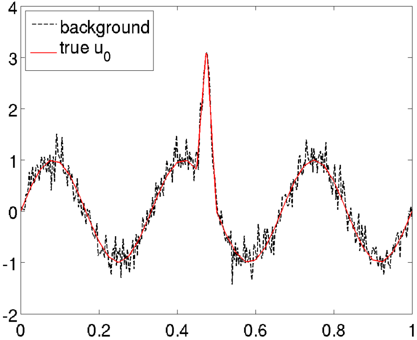


η_j

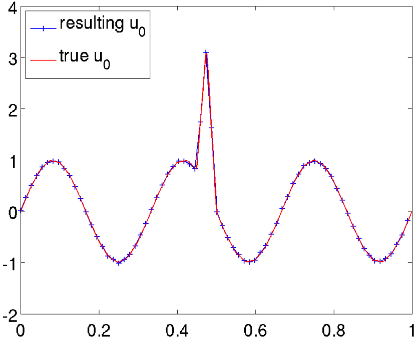


$$\epsilon_j = w_j \left\langle \Delta\lambda_{i+1}|_j, [(\tilde{R}_{i+1} + \tilde{H}_{i+1} B \tilde{H}_{i+1}^T)\Delta\lambda_{i+1}]|_j \right\rangle$$

Example 2: Control variable



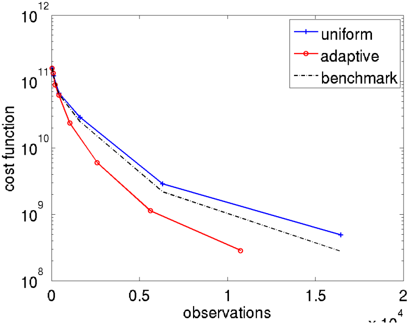
Background vector and true $u(0)$



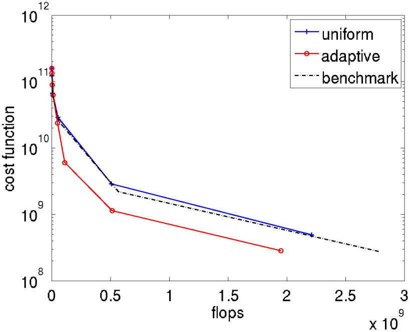
Algorithm solution and true $u(0)$

Example 2: Cost function and RMS error

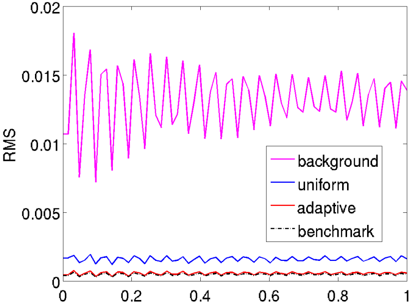
Cost f. VS nb obs



Cost f. VS flops



RMS error VS time



Bibliography

- **Borzi A. and Schulz V.**: Multigrid methods for PDE optimization, *SIAM Rev.*, 51 (2), 361-395, 2009.
- **Brandt A.**: Guide to a multigrid development, *Multigrid Methods, Lecture Notes in Mathematics*, Hackbush, Trottenber (eds), Springer Berlin Heidelberg, 960, 220-312, 1982.
- **Cioca A., Sandu A., de Sturler E.**: Efficient methods for computing observation impact in 4D-Var data assimilation, *Comput. Geosci.*, 17, 975-990, 2013.
- **Debreu L., Neveu E., Simon E., Le Dimet F.-X., Vidard A.**: Multigrid solvers and multigrid preconditioners for the solution of variational data assimilation problems, In revision.
- **Gratton S., Rincon-Camacho M., Simon E. and Toint P.**: Observations thinning in data assimilation computations, *EURO Journal on Computational Optimization*, 3, 31-51, 2015.
- **Gratton S. and Tshimanga J.**: An observation-space formulation of variational assimilation using a restricted preconditioned conjugates gradient algorithm, *Quarterly Journal of the Royal Meteorological Society*, 135, 1573-1585, 2009.
- **Lewis R.M. and Nash S.G.**: Model problems for the multigrid optimization of systems governed by differential equations, *SIAM J. Sci. Comput.*, 26 (6), 1811-1837, 2005.
- **Neveu E., Debreu L., Le Dimet F.-X.**: Multigrid methods and data assimilation: Convergence study and first experiments on nonlinear equations, *ARIMA*, 14, 63-80, 2011.