Some challenges of four-dimensional data assimilation problems

Juan Carlos De los Reyes

Centro de Modelización Matemática (MODEMAT)
Escuela Politécnica Nacional de Ecuador

New trends in PDE constrained optimization
Linz, 2019

Contenidos

ModeMat

Outline

Motivation

Data assimilation in Numerical Weather Prediction

- Data assimilation methods aim at finding a good initial condition of the athmospheric system in order to get better weather forecasts;

Motivation

Data assimilation in Numerical Weather Prediction

- Data assimilation methods aim at finding a good initial condition of the athmospheric system in order to get better weather forecasts;
- Information can be obtained mainly from ground stations, radionsonds or satellite images;

Motivation

Data assimilation in Numerical Weather Prediction

- Data assimilation methods aim at finding a good initial condition of the athmospheric system in order to get better weather forecasts;
- Information can be obtained mainly from ground stations, radionsonds or satellite images;
- Reconstruction results depend strongly on the number of observations, which can be very limited in some cases.



Equations of the athmosphere

Basic model

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial y} - w\frac{\partial u}{\partial z} + \frac{uv\tan(\phi)}{a} - \frac{uw}{a} - \frac{1}{\rho}\frac{\partial p}{\partial x} - 2\Omega(w\cos(\phi) - v\sin(\phi)) + Fr_x$$

$$\frac{\partial v}{\partial t} = -u\frac{\partial v}{\partial x} - v\frac{\partial v}{\partial y} - w\frac{\partial v}{\partial z} + \frac{u^2\tan(\phi)}{a} - \frac{uw}{a} - \frac{1}{\rho}\frac{\partial p}{\partial y} - 2\Omega u\sin(\phi) + Fr_y$$

$$\frac{\partial w}{\partial t} = -u\frac{\partial w}{\partial x} - v\frac{\partial w}{\partial y} - w\frac{\partial w}{\partial z} + \frac{u^2 + v^2}{a} - \frac{1}{\rho}\frac{\partial p}{\partial z} + 2\Omega u\cos(\phi) - g + Fr_z$$

$$\frac{\partial T}{\partial t} = -u\frac{\partial T}{\partial x} - v\frac{\partial T}{\partial y} + (\gamma - \gamma_d)w + \frac{1}{c_p}\frac{dH}{dt}$$

$$\frac{\partial \rho}{\partial t} = -u\frac{\partial \rho}{\partial x} - v\frac{\partial \rho}{\partial y} - w\frac{\partial \rho}{\partial z} - \rho\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right)$$

$$\frac{\partial q}{\partial t} = -u\frac{\partial q}{\partial x} - v\frac{\partial q}{\partial y} - w\frac{\partial q}{\partial z} + Q_v$$

$+$ Boundary conditions

$+$ Initial conditions

Equations of the athmosphere

Basic model

Mode Mat

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial y} - w\frac{\partial u}{\partial z} + \frac{uv\tan(\phi)}{a} - \frac{uw}{a} - \frac{1}{\rho}\frac{\partial p}{\partial x} - 2\Omega(w\cos(\phi) - v\sin(\phi)) + Fr_x$$

$$\frac{\partial v}{\partial t} = -u\frac{\partial v}{\partial x} - v\frac{\partial v}{\partial y} - w\frac{\partial v}{\partial z} + \frac{u^2\tan(\phi)}{a} - \frac{uw}{a} - \frac{1}{\rho}\frac{\partial p}{\partial y} - 2\Omega u\sin(\phi) + Fr_y$$

$$\frac{\partial w}{\partial t} = -u\frac{\partial w}{\partial x} - v\frac{\partial w}{\partial y} - w\frac{\partial w}{\partial z} + \frac{u^2 + v^2}{a} - \frac{1}{\rho}\frac{\partial p}{\partial z} + 2\Omega u\cos(\phi) - g + Fr_z$$

$$\frac{\partial T}{\partial t} = -u\frac{\partial T}{\partial x} - v\frac{\partial T}{\partial y} + (\gamma - \gamma_d)w + \frac{1}{c_p}\frac{dH}{dt}$$

$$\frac{\partial \rho}{\partial t} = -u\frac{\partial \rho}{\partial x} - v\frac{\partial \rho}{\partial y} - w\frac{\partial \rho}{\partial z} - \rho\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right)$$

$$\frac{\partial q}{\partial t} = -u\frac{\partial q}{\partial x} - v\frac{\partial q}{\partial y} - w\frac{\partial q}{\partial z} + Q_v$$

$+$ Boundary conditions

$+$ Initial conditions

Approaches to data assimilation

Approaches based on static covariance matrices

- Optimal interpolation
- 3D-Var

Recent approaches based on dynamic covariance matrices

- 4D-Var
- Extended Kalman filters
- Ensemble methods

Eugenia Kalnay
Atmospheric Modeling, Data Assimilation and Predictability
Cambridge Univ. Press, 2002.

Optimal interpolation

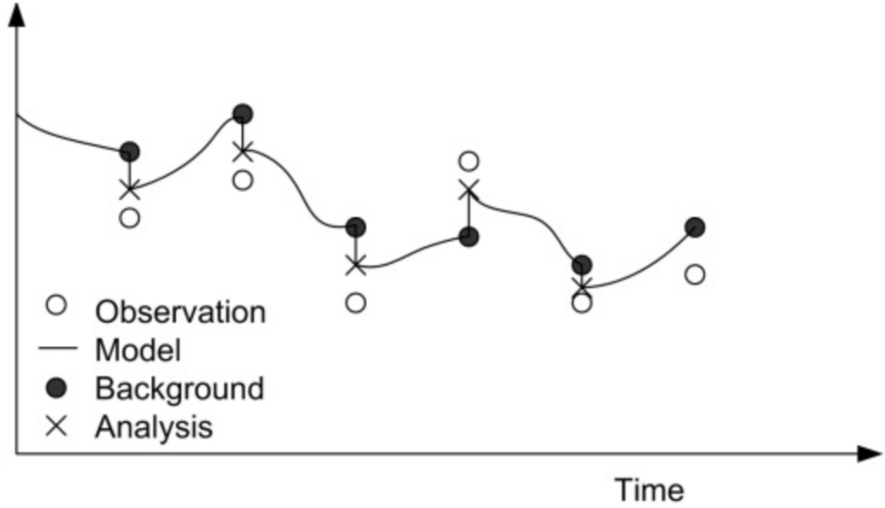
- u_b : background information vector
- u_a : “analysis” (estimation of the state)
- z : observation vector
- $z = Hu + v$, where H is an observation operator and v is the observation error
- all variables are assumed to be Gaussian

Kalman filter

$$u_a = u_b + K(z - Hu_b),$$

where $K = BH^T(HBH^T + R)^{-1}$ is the *Kalman gain* corresponding to the linear unbiased estimator of minimim variance.

Intuition of OI



3D-Var

MAP estimation

- The data assimilation problem may be treated via a Bayesian approach to find the maximum-a-posteriori (MAP) estimator.

3D-Var

MAP estimation

- The data assimilation problem may be treated via a Bayesian approach to find the maximum-a-posteriori (MAP) estimator.
- Using the Gaussian probability density functions

$$
\begin{aligned}
p(z|u) &\propto \exp\left(-\frac{1}{2}(z - Hu)^T R^{-1}(z - Hu)\right) \\
p(u) &\propto \exp\left(-\frac{1}{2}(u_b - u)^T B^{-1}(u_b - u)\right)
\end{aligned}
$$

3D-Var

MAP estimation

- The data assimilation problem may be treated via a Bayesian approach to find the maximum-a-posteriori (MAP) estimator.
- Using the Gaussian probability density functions

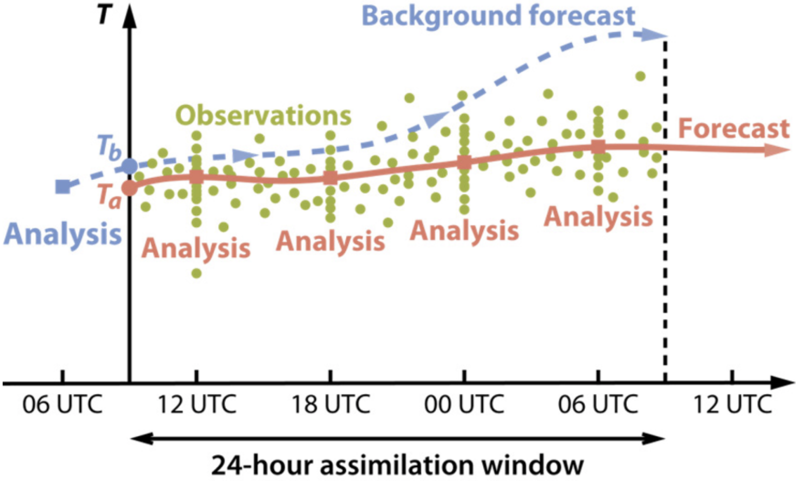
$$p(z|u) \propto \exp\left(-\frac{1}{2}(z - Hu)^T R^{-1}(z - Hu)\right)$$

$$p(u) \propto \exp\left(-\frac{1}{2}(u_b - u)^T B^{-1}(u_b - u)\right)$$

- Using Bayes formula, $p(u|z) = \frac{p(z|u)p(u)}{p(z)}$, the MAP estimator corresponds to the solution of the 3D-Var problem

$$\min \frac{1}{2}(z - Hu)^T R^{-1}(z - Hu) + \frac{1}{2}(u_b - u)^T B^{-1}(u_b - u).$$

Intuition of 4D-Var



4D-Var

Le Dimet and Talagrand (1986)

Finite dimensional problem

$$\begin{aligned}\min_u J(y,u) &= \frac{1}{2}\sum_{i=0}^{l}\left[H(y(t_i)) - z_o(t_i)\right]^T R_i^{-1}\left[H(y(t_i)) - z_o(t_i)\right] \\ &\quad + \frac{1}{2}\left[u-u_b\right]^T B^{-1}\left[u-u_b\right]\end{aligned}$$

subject to:

$$\begin{aligned} y(t_l) &= M(y(t_0)) && \text{(Dynamical system)} \\ y(t_0) &= u && \text{(Initial condition).}\end{aligned}$$

4D-Var

Le Dimet and Talagrand (1986)

Finite dimensional problem

$$
\begin{aligned}
\min_u J(y,u) &= \frac{1}{2}\sum_{i=0}^{l}\left[H(y(t_i)) - z_o(t_i)\right]^T R_i^{-1}\left[H(y(t_i)) - z_o(t_i)\right] \\
&\quad + \frac{1}{2}\left[u-u_b\right]^T B^{-1}\left[u-u_b\right]
\end{aligned}
$$

subject to:

$$
\begin{aligned}
y(t_l) &= M(y(t_0)) && \text{(Dynamical system)} \\
y(t_0) &= u && \text{(Initial condition).}
\end{aligned}
$$

Features

- The dynamic problem incorporates all observations in a given time window;
- The nonlinear dynamics may be taken into account;
- The operational use is still a computational challenge.

4D-Var

Le Dimet and Talagrand (1986)

Finite dimensional problem

$$\begin{aligned}
\min_u J(y,u) &= \frac{1}{2}\sum_{i=0}^{l}\left[H(y(t_i)) - z_o(t_i)\right]^T R_i^{-1}\left[H(y(t_i)) - z_o(t_i)\right] \\
&\quad + \frac{1}{2}\left[u-u_b\right]^T B^{-1}\left[u-u_b\right]
\end{aligned}$$

subject to:

$$\begin{aligned}
y(t_l) &= M(y(t_0)) && \text{(Dynamical system)} \\
y(t_0) &= u && \text{(Initial condition).}
\end{aligned}$$

What about the infinite-dimensional problem?

Outline

Semilinear problem

ModeMat

$$\min_u J(y,u) = \frac{1}{2}\sum_{k,i}[y(x_k,t_i) - z_0(x_k,t_i)]^2 + \frac{1}{2}\|u-u_b\|^2_{B^{-1}}$$

subject to:

$$\begin{aligned} \frac{\partial y}{\partial t} + Ay + g(y) &= 0 && \text{in } Q = \Omega \times]0,T[\\ y &= 0 && \text{on } \Sigma = \Gamma \times]0,T[\\ y(x,0) &= u && \text{in } \Omega, \end{aligned}$$

Difficulties

- Higher regularity of the initial condition is required to get some sort of continuity of the state.
- Pointwise misfits in the cost leads to right hand sides in $\mathcal{M}(Q)$ for the adjoint equation. Ill-posedness!

Semilinear problem

$$\min_u J(y,u) = \frac{1}{2}\sum_{k,i}[y(x_k,t_i) - z_0(x_k,t_i)]^2 + \frac{1}{2}\|u-u_b\|^2_{B^{-1}}$$

subject to:

$$\begin{aligned} \frac{\partial y}{\partial t} + Ay + g(y) &= 0 \quad \text{in } Q = \Omega\times]0,T[\\ y &= 0 \quad \text{on } \Sigma = \Gamma\times]0,T[\\ y(x,0) &= u \quad \text{in } \Omega, \end{aligned}$$

Difficulties

- Higher regularity of the initial condition is required to get some sort of continuity of the state.
- Pointwise misfits in the cost leads to right hand sides in $\mathcal{M}(Q)$ for the adjoint equation. Ill-posedness!

Semilinear problem

$$\min_u J(y,u) = \frac{1}{2}\int_0^T \sum_{k,i} w_k \sigma_i \rho_i(t)[y(x_k,t) - z_o(x_k,t)]^2\, dt$$

$$+ \frac{1}{2}\|u - u_b\|^2_{B^{-1}} + \frac{\vartheta}{2}\|\nabla(u - u_b)\|^2_{L^2(\Omega)}$$

subject to:

$$\begin{aligned} \frac{\partial y}{\partial t} + Ay + g(y) &= 0 \quad \text{in } Q = \Omega\times]0,T[\\ y &= 0 \quad \text{on } \Sigma = \Gamma\times]0,T[\\ y(x,0) &= u \quad \text{in } \Omega, \end{aligned}$$

Difficulties

- Higher regularity of the initial condition is required to get some sort of continuity of the state.
- Pointwise misfits in the cost leads to right hand sides in $\mathcal{M}(Q)$ for the adjoint equation. Ill-posedness!
- w and σ are binary vectors, and $\rho_i(t)$ support functions

Well-posedness

Assumption on the nonlinearity

- $g = g(x,t,y) : Q \times \mathbb{R} \mapsto \mathbb{R}$ satisfies the Carathéodory conditions and is uniformly bounded at the origin, i.e., $|g(x,t,0)| \leq K$, for some $K > 0$,
- g is monotone increasing with respect to y for almost every $(x,t) \in Q$,
- g is twice continuously differentiable with respect to y and

$$|g_y(x,t,y)| + |g_{yy}(x,t,y)| \leq K,$$

for some $K > 0$, for almost all $(x,t) \in Q$ and any $y \in \mathbb{R}$.

Theorem

If $u \in H_0^1(\Omega)$ and the nonlinear term verifies the assumption, then the semilinear equation has a unique solution $y \in H^{2,1}(Q) \hookrightarrow L^2(0,T;C(\Omega))$. Moreover,

$$\|y\|_{H^{2,1}(Q)} \leq c(1 + \|u\|_{H_0^1(\Omega)}), \text{ for some } c > 0.$$

Differentiability

Mode Mat

Theorem

The control-to-state mapping $S : H_0^1(\Omega) \to H^{2,1}(Q)$, $u \mapsto S(u) = y$, is Gâteaux differentiable. Its derivative, in direction $h \in H_0^1(\Omega)$, is given by $\eta \in H^{2,1}(Q)$ solution of:

$$\begin{cases} \frac{\partial \eta}{\partial t} + A\eta + g'(y)\eta = & 0 & \text{in } Q \\ \eta = & 0 & \text{on } \Sigma \\ \eta(x,0) = & h & \text{in } \Omega, \end{cases}$$

Adjoint equation

$$\begin{aligned} -\frac{\partial p}{\partial t} + A^* p + g'(y)p = & \quad \mu & \text{in } Q \\ p = & \quad 0 & \text{on } \Sigma \\ p(x,T) = & \quad 0 & \text{in } \Omega, \end{aligned}$$

with $\mu \in L^2(0,T;\mathcal{M}(\Omega))$ has a unique solution $p \in L^2(0,T;W_0^{1,r}(\Omega))$, with $r \in [1, \frac{m}{m-1}[$. Casas-Clason-Kunisch 2013, Meyer-Susu 2017

Optimality system

ModeMat

State equation (in strong form):

$$\begin{aligned} \frac{\partial \bar{y}}{\partial t} + A\bar{y} + g(\bar{y}) &= 0 \quad \text{in } Q, \\ \bar{y} &= 0 \quad \text{on } \Sigma, \\ \bar{y}(x,0) &= \bar{u} \quad \text{in } \Omega. \end{aligned}$$

Adjoint equation (in very weak form):

$$\begin{aligned} -\frac{\partial \bar{p}}{\partial t} + A^*\bar{p} + g'(\bar{y})\bar{p} &= \sum_{k,i} w_k \sigma_i \rho_i(t) \left[\bar{y}(x,t) - z_o(x,t)\right] \otimes \delta(x - x_k) \quad \text{in } Q, \\ \bar{p} &= 0 \quad \text{on } \Sigma, \\ \bar{p}(x,T) &= 0 \quad \text{in } \Omega. \end{aligned}$$

Gradient equation (in weak form):

$$\begin{aligned} -\vartheta\Delta(\bar{u} - u_b) + B^{-1}(\bar{u} - u_b) + \bar{p}(0) &= 0 \quad \text{in } \Omega, \\ \bar{u} &= 0 \quad \text{on } \Gamma. \end{aligned}$$

Extensions

- Interplay between function spaces and type of observations;

Extensions

- Interplay between function spaces and type of observations;
- Solutions for the dynamics of the atmosphere are not necessarily continuous;

Extensions

- Interplay between function spaces and type of observations;
- Solutions for the dynamics of the atmosphere are not necessarily continuous;
- Variational DA models typically contain a classical Tikhonov regularization $\|u - u_b\|^2_{B^{-1}}$, which is not appropriate to reconstruct *sharp fronts*

Extensions

- Interplay between function spaces and type of observations;
- Solutions for the dynamics of the atmosphere are not necessarily continuous;
- Variational DA models typically contain a classical Tikhonov regularization $\|u - u_b\|^2_{B^{-1}}$, which is not appropriate to reconstruct *sharp fronts*
- Recently Freitag and co. (2013, 2015) proposed an alternative total variation (TV) regularization to recover solutions with sharp fronts.

Extensions

- Interplay between function spaces and type of observations;
- Solutions for the dynamics of the atmosphere are not necessarily continuous;
- Variational DA models typically contain a classical Tikhonov regularization $\|u - u_b\|^2_{B^{-1}}$, which is not appropriate to reconstruct *sharp fronts*
- Recently Freitag and co. (2013, 2015) proposed an alternative total variation (TV) regularization to recover solutions with sharp fronts.
- A drawback of TV is, however, the staircase effect, which may lead to undesired artifacts in the reconstruction.

Extensions

- Interplay between function spaces and type of observations;
- Solutions for the dynamics of the atmosphere are not necessarily continuous;
- Variational DA models typically contain a classical Tikhonov regularization $\|u - u_b\|_{B^{-1}}^2$, which is not appropriate to reconstruct *sharp fronts*
- Recently Freitag and co. (2013, 2015) proposed an alternative total variation (TV) regularization to recover solutions with sharp fronts.
- A drawback of TV is, however, the staircase effect, which may lead to undesired artifacts in the reconstruction.
- We studied second order total generalized variation regularization De los Reyes, Loayza (2019)

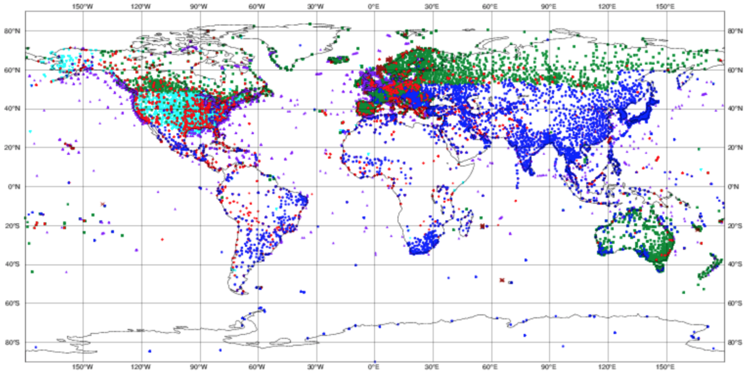
Extensions

- Interplay between function spaces and type of observations;
- Solutions for the dynamics of the atmosphere are not necessarily continuous;
- Variational DA models typically contain a classical Tikhonov regularization $\|u - u_b\|_{B^{-1}}^2$, which is not appropriate to reconstruct *sharp fronts*
- Recently Freitag and co. (2013, 2015) proposed an alternative total variation (TV) regularization to recover solutions with sharp fronts.
- A drawback of TV is, however, the staircase effect, which may lead to undesired artifacts in the reconstruction.
- We studied second order total generalized variation regularization De los Reyes, Loayza (2019)

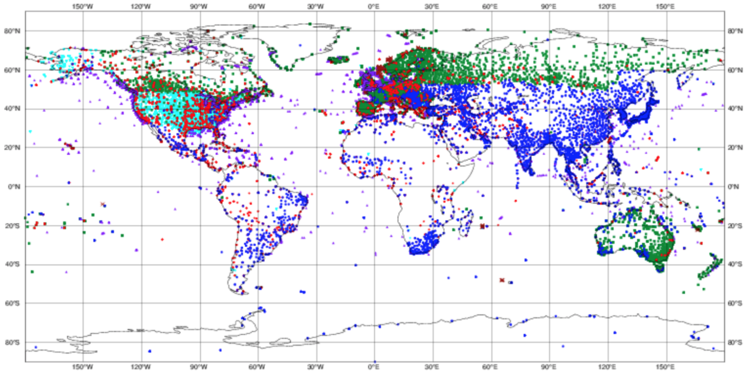
Is this of operational use?

Outline

Observation stations

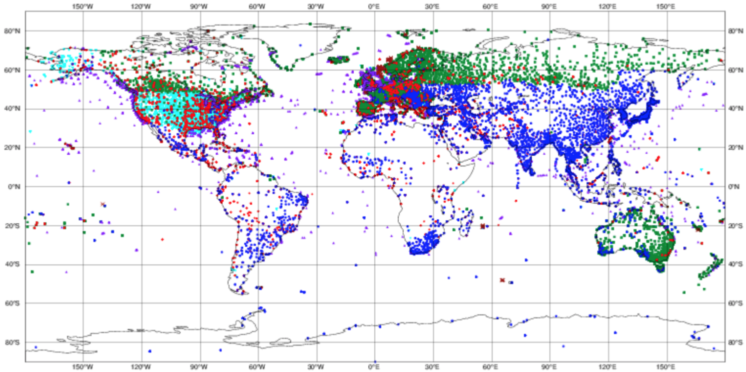


Observation stations



More observations are required to further improve forecasts

Observation stations



Problem

One has to make the best possible decision about where to locate the next observation stations to obtain better reconstructions of the initial condition.

State-of-the-art

- Optimal filtering Bensoussan and collab. (1971, 1972), Burns and collab. (1994, 1995, 1998, 2009, 2011), Rautenberg and collab. (2015, 2016)
- Observability approaches Privat, Trelat, Zuazua (2013, 2015, 2017)
- Supervised learning Haber, Horesh, Tenorio (2008, 2010), De los Reyes, Schönlieb and collab. (2013, 2016, 2017), Kunisch and collab. (2013, 2018)
- Bayesian optimal experimental design Alexanderian et al. (2014, 2015), Herzog et al. (2015, 2017)

State-of-the-art

- Optimal filtering Bensoussan and collab. (1971, 1972), Burns and collab. (1994, 1995, 1998, 2009, 2011), Rautenberg and collab. (2015, 2016)
- Observability approaches Privat, Trelat, Zuazua (2013, 2015, 2017)
- Supervised learning Haber, Horesh, Tenorio (2008, 2010), De los Reyes, Schönlieb and collab. (2013, 2016, 2017), Kunisch and collab. (2013, 2018)
- Bayesian optimal experimental design Alexanderian et al. (2014, 2015), Herzog et al. (2015, 2017)

Features and goals

- Bilevel learning approach: flexibility about quality measures and more natural framework for nonlinear lower level problems.
- Analysis of the resulting problem in function space: optimality conditions and multiplier regularity.
- Design of efficient solution algorithms.

Bilevel learning problem

Mixed integer-infinite dimensional optimization

$$\min_{w,\sigma\in\{0,1\}} \iint_Q \sum_{j=1}^{N} L(y_j, y_j^\dagger)\, dxdt + \beta \int_\Omega \sum_{j=1}^{N} l(u_j, u_j^\dagger)\, dx + \beta_w \sum_k w_k + \beta_\sigma \sum_i \sigma_i$$

subject to $(\forall j = 1, \dots, N)$:

$$\begin{cases} \min_{u_j} \frac{1}{2} \int_0^T \sum_{k,i} \left(w_k \sigma_i \rho_i(t) [y_j(x_k, t) - z_{oj}(x_k, t)]^2 \right) dt \\ \qquad\qquad + \frac{1}{2} \|u_j - u_{bj}\|_{B^{-1}} + \frac{\vartheta}{2} \|\nabla(u_j - u_{bj})\|^2_{L^2(\Omega)} \\ \text{subject to:} \\ \frac{\partial y_j}{\partial t} + Ay_j + g(y_j) = 0 \quad \text{in } Q \\ y_j = 0 \quad \text{on } \Sigma \\ y_j(0) = u_j \quad \text{in } \Omega. \end{cases}$$

Bilevel learning problem

Mixed integer-infinite dimensional optimization

$$\min_{w,\sigma\in\{0,1\}} \iint_Q \sum_{j=1}^{N} L(y_j, y_j^\dagger)\, dxdt + \beta \int_\Omega \sum_{j=1}^{N} l(u_j, u_j^\dagger)\, dx + \beta_w \sum_k w_k + \beta_\sigma \sum_i \sigma_i$$

- $\beta_w > 0$ corresponds to the sparsity penalty term for w and $\beta_\sigma > 0$ the one for σ;
- The training set $\left(u_j^\dagger, y_j^\dagger\right)$, $\forall j = 1, \ldots, N$ is built from improved reconstructions of the initial condition and the observed state;
- We aim at learning the vector of placements w such that the quality measure of the training set is minimized in average $\left(u_j^\dagger, y_j^\dagger\right)$.

Dealing with sparsity

- Relaxing the integer constraints, $0 \leq w, \sigma \leq 1$, we simply get linearity with respect to w and σ:

$$\min_{0\leq w,\sigma\leq 1} \iint_Q \sum_{j=1}^{N} L(y_j, y_j^\dagger)\, dxdt + \beta \int_\Omega \sum_{j=1}^{N} l(u_j, u_j^\dagger)\, dx + \beta_w \sum_k w_k + \beta_\sigma \sum_i \sigma_i$$

Dealing with sparsity

- Relaxing the integer constraints, $0 \le w, \sigma \le 1$, we simply get linearity with respect to w and σ:

$$\min_{0 \le w, \sigma \le 1} \iint_Q \sum_{j=1}^{N} L(y_j, y_j^\dagger)\, dxdt + \beta \int_\Omega \sum_{j=1}^{N} l(u_j, u_j^\dagger)\, dx + \beta_w \sum_k w_k + \beta_\sigma \sum_i \sigma_i$$

- Although sparsity is obtained as a result of the ℓ_1-norm penalty, the number of values different from 0 or 1 makes the solution difficult to interprete

Dealing with sparsity

- Relaxing the integer constraints, $0 \leq w, \sigma \leq 1$, we simply get linearity with respect to w and σ:

$$\min_{0\leq w,\sigma\leq 1} \iint_Q \sum_{j=1}^{N} L(y_j, y_j^\dagger)\, dxdt + \beta \int_\Omega \sum_{j=1}^{N} l(u_j, u_j^\dagger)\, dx + \beta_w \sum_k w_k + \beta_\sigma \sum_i \sigma_i$$

- Although sparsity is obtained as a result of the ℓ_1-norm penalty, the number of values different from 0 or 1 makes the solution difficult to interprete
- We consider also an approximation of the l_p norm, with $p \in (0, 1)$. For instance

$$\phi(x) = \begin{cases} \frac{x}{\epsilon} & \text{if } 0 \leq x < \frac{\epsilon}{2} \\ p_\epsilon(x) & \text{if } \frac{\epsilon}{2} < x \leq 2\epsilon \\ 1 & \text{if } 2\epsilon \leq x \leq 1 \end{cases}$$

Alexandarian et al. 2014

Control of a singular system with measures

Replacing the lower level problems by their necessary optimality condition:

$$\min_{0\le w,\sigma\le 1} J(y,p,u,w) = \iint_Q \sum_{j=1}^{N} L(y_j, y_j^\dagger)\, dxdt + \beta \int_\Omega \sum_{j=1}^{N} l(u_j, u_j^\dagger)\, dx + \beta_w \sum_k w_k + \beta_\sigma \sum_i \sigma_i$$

subject to:

$$\begin{aligned}
\frac{\partial y_j}{\partial t} + Ay_j + g(y_j) &= 0\\
y_j|_\Gamma &= 0\\
y_j(0) &= u_j\\
-\frac{\partial p_j}{\partial t} + A^* p_j + g'(y_j)p_j &= \sum_{k,i} w_k \sigma_i \rho_i(t)\, [\bar{y}(x,t) - z_o(x,t)] \otimes \delta(x - x_k)\\
p_j|_\Gamma &= 0\\
p_j(T) &= 0\\
-\vartheta\Delta(u_j - u_b) + B^{-1}(u_j - u_b) &= -p_j(0)\\
\bar{u}|_\Gamma &= 0.
\end{aligned}$$

Control of a singular system with measures

Replacing the lower level problems by their necessary optimality condition:

$$\min_{0\le w,\sigma\le 1} J(y,p,u,w) = \iint_Q \sum_{j=1}^{N} L(y_j, y_j^\dagger)\ dxdt + \beta \int_\Omega \sum_{j=1}^{N} l(u_j, u_j^\dagger)\ dx + \beta_w \sum_k w_k + \beta_\sigma \sum_i \sigma_i$$

subject to:

$$\begin{aligned}
\frac{\partial y_j}{\partial t} + Ay_j + g(y_j) &= 0\\
y_j|_\Gamma &= 0\\
y_j(0) &= u_j\\
-\frac{\partial p_j}{\partial t} + A^* p_j + g'(y_j)p_j &= \sum_{k,i} w_k \sigma_i \rho_i(t)\, [\bar{y}(x,t) - z_o(x,t)] \otimes \delta(x - x_k)\\
p_j|_\Gamma &= 0\\
p_j(T) &= 0\\
-\vartheta\Delta(u_j - u_b) + B^{-1}(u_j - u_b) &= -p_j(0)\\
\bar{u}|_\Gamma &= 0.
\end{aligned}$$

Casas, Clason, Kunisch, Neitzel, Pieper, Vexler, ...

Control of a singular system with measures

Mode Mat MMM

Replacing the lower level problems by their necessary optimality condition:

$$\min_{0\leq w,\sigma\leq 1} J(y,p,u,w) = \iint_Q \sum_{j=1}^N L(y_j, y_j^\dagger)\, dxdt + \beta \int_\Omega \sum_{j=1}^N l(u_j, u_j^\dagger)\, dx + \beta_w \sum_k w_k + \beta_\sigma \sum_i \sigma_i$$

subject to:

$$\begin{aligned}
\frac{\partial y_j}{\partial t} + Ay_j + g(y_j) &= 0\\
y_j|_\Gamma &= 0\\
y_j(0) &= u_j\\
-\frac{\partial p_j}{\partial t} + A^* p_j + g'(y_j)p_j &= \sum_{k,i} w_k \sigma_i \rho_i(t)\,[\bar{y}(x,t) - z_o(x,t)] \otimes \delta(x - x_k)\\
p_j|_\Gamma &= 0\\
p_j(T) &= 0\\
-\vartheta\Delta(u_j - u_b) + B^{-1}(u_j - u_b) &= -p_j(0)\\
\bar{u}|_\Gamma &= 0.
\end{aligned}$$

Difficulty: No unique solution of the optimality system!

Adapted penalty approach

For $(\bar{y}, \bar{p}, \bar{u}) \in H^{2,1}(Q) \times L^2(W_0^{1,r}(\Omega)) \times H_0^1(\Omega)$ solution of the bilevel problem, find $(y_\gamma, p_\gamma, \mathbf{w}_\gamma, s_\gamma, v_\gamma) \in H^{2,1}(Q) \times L^2(W_0^{1,r}(\Omega)) \times V_{ad} \times L^2(Q) \times L^2(Q)$ that solves:

$$\min_{\mathbf{w},s,v} J_\gamma(y,p,\mathbf{w},s,v) = J(y,p,\mathbf{w}) + \frac{\gamma}{2} \iint_Q (s - g(y))^2 + \frac{\gamma}{2} \iint_Q (v - g'(y)p)^2$$
$$+ \frac{1}{2} \|p - \bar{p}\|^2_{L^2(Q)} + \frac{1}{2} \iint_Q (s - g(\bar{y}))^2 + \frac{1}{2} \iint_Q (v - g'(\bar{y})\bar{p})^2$$

subject to:

$$\begin{aligned} \frac{\partial y}{\partial t} + Ay + s &= 0 && \text{in } Q \\ y &= 0 && \text{on } \Sigma \\ y(0) &= -G^{-1}p(0) && \text{in } \Omega \end{aligned}$$

$$\begin{aligned} -\frac{\partial p}{\partial t} + A^* p + v &= \sum_{k,i} w_k \sigma_i \rho_i(t) \left[y(x,t) - z_o(x,t) \right] \otimes \delta(x - x_k) && \text{in } Q \\ p &= 0 && \text{on } \Sigma \\ p(T) &= 0 && \text{in } \Omega. \end{aligned}$$

$$\langle Gu, \tau \rangle_{H^{-1},H_0^1} := \int_\Omega (u - u_b) B^{-1} \tau + \vartheta \int_\Omega \nabla(u - u_b).\nabla \tau = -\int_\Omega p(0)\tau, \forall \tau \in H_0^1(\Omega).$$

Consistency

The penalized problem has at least one solution $(y_\gamma, p_\gamma, \mathbf{w}_\gamma, s_\gamma, v_\gamma) \in H^{2,1}(Q) \times L^2(W_0^{1,r}(\Omega)) \times V_{ad} \times L^2(Q) \times L^2(Q)$ and corresponding Lagrange multipliers $(\eta_\gamma, \zeta_\gamma) \in L^2(Q) \times H^{2,1}(Q)$. Moreover, the sequence $\{(y_\gamma, p_\gamma, \mathbf{w}_\gamma, s_\gamma, v_\gamma)\}$ converges strongly to the solution $(\bar{y}, \bar{p}, \bar{\mathbf{w}}, g(\bar{y}), g'(\bar{y})\bar{p})$.

- Structure of penalized cost and properties of state and adjoint eq. to get boundedness of variables;
- Compact embedding $H^{2,1}(Q) \hookrightarrow\hookrightarrow L^\mu(Q)$, with $\mu \leq 10$, and continuous embedding $W^{1,r}(\Omega) \hookrightarrow L^q(\Omega)$ for $q \leq \frac{mr}{m-r}$;
- Existence of multipliers using the linearity of the constraints.
- Using Hölder properties of Bochner spaces and properties of the PDE, we get uniform bounds on the regularized variables.

Consistency

The penalized problem has at least one solution $(y_\gamma, p_\gamma, \mathbf{w}_\gamma, s_\gamma, v_\gamma) \in H^{2,1}(Q) \times L^2(W_0^{1,r}(\Omega)) \times V_{ad} \times L^2(Q) \times L^2(Q)$ and corresponding Lagrange multipliers $(\eta_\gamma, \zeta_\gamma) \in L^2(Q) \times H^{2,1}(Q)$. Moreover, the sequence $\{(y_\gamma, p_\gamma, \mathbf{w}_\gamma, s_\gamma, v_\gamma)\}$ converges strongly to the solution $(\bar{y}, \bar{p}, \bar{\mathbf{w}}, g(\bar{y}), g'(\bar{y})\bar{p})$.

- Structure of penalized cost and properties of state and adjoint eq. to get boundedness of variables;
- Compact embedding $H^{2,1}(Q) \hookrightarrow\hookrightarrow L^\mu(Q)$, with $\mu \leq 10$, and continuous embedding $W^{1,r}(\Omega) \hookrightarrow L^q(\Omega)$ for $q \leq \frac{mr}{m-r}$;
- Existence of multipliers using the linearity of the constraints.
- Using Hölder properties of Bochner spaces and properties of the PDE, we get uniform bounds on the regularized variables.

Lagrange multipliers

There exists an adjoint state $(\eta_j, \zeta_j, \tau_j) \in L^2(Q) \times H^{2,1}(Q) \times H_0^1(\Omega)$, for all $j = 1, \ldots, N$, and KKT multipliers $\lambda^a, \lambda^b \in \mathbb{R}^{n_s} \times \mathbb{R}^{n_T}$.

Adjoint system of the bilevel problem

ModeMat

$$
\begin{aligned}
-\frac{\partial \eta_j}{\partial t} + A^*\eta_j + g'(y_j)\eta_j + g''(y_j)p_j\zeta_j & \\
= \sum_{k,i} w_k \sigma_i \rho_i(t)\zeta_j(x,t) \otimes \delta(x - x_k) - \nabla_{y_j} L(y_j) & \quad \text{in } Q \\
\eta_j = & 0 \quad \text{on } \Sigma \\
\eta_j(T) = & 0 \quad \text{in } \Omega.
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \zeta_j}{\partial t} + A\zeta_j + g'(y_j)\zeta_j = & 0 \quad \text{in } Q \\
\zeta_j = & 0 \quad \text{on } \Sigma \\
\zeta_j(0) = & -\tau_j \quad \text{in } \Omega
\end{aligned}
$$

$$
\begin{aligned}
-\vartheta\Delta\tau_j + B^{-1}\tau_j = & \eta_j(0) - \beta\nabla_{u_j} l(u_j) \quad \text{in } \Omega \\
\tau_j = & 0 \quad \text{on } \Gamma,
\end{aligned}
$$

very weakly, strongly and weakly, respectively.

Karush-Kuhn-Tucker condition

Mode Mat

Gradient system:

$$\beta_w - \sum_{j=1}^{N} \int_0^T \sum_i \sigma_i \rho_i(t) \zeta_j(x_k,t) \left(y_j(x_k,t) - z_{oj}(x_k,t)\right) \, dt = \lambda_k^a - \lambda_k^b, \quad \forall k,$$

$$\beta_\sigma - \sum_{j=1}^{N} \int_0^T \sum_k w_k \rho_i(t) \zeta_j(x_k,t) \left(y_j(x_k,t) - z_{oj}(x_k,t)\right) \, dt = \lambda_{n_s+i}^a - \lambda_{n_s+i}^b, \quad \forall i.$$

Complementarity system:

$$\begin{aligned}
&\lambda_r^a \geq 0, \lambda_r^b \geq 0, && \text{for all } r = 1, \ldots, n_s + n_T \\
&\lambda_k^a \bar{w}_k = \lambda_k^b (\bar{w}_k - 1) = 0, && \text{for all } k = 1, \ldots, n_s \\
&\lambda_{n_s+i}^a \bar{\sigma}_i = \lambda_{n_s+i}^b (\bar{\sigma}_i - 1) = 0, && \text{for all } i = 1, \ldots, n_T \\
&0 \leq \bar{w}_k \leq 1, && \text{for all } k = 1, \ldots, n_s \\
&0 \leq \bar{\sigma}_i \leq 1, && \text{for all } i = 1, \ldots, n_T.
\end{aligned}$$

Outline

Projected BFGS method

- Let S be an index set and R_S the matrix defined by

$$R_S = \begin{cases} \delta_{ij}, & \text{if } i \in S \text{ or } j \in S; \\ 0, & \text{if not.} \end{cases}$$

.

Projected BFGS method

- Let S be an index set and R_S the matrix defined by

$$R_S = \begin{cases} \delta_{ij}, & \text{if } i \in S \text{ or } j \in S; \\ 0, & \text{if not.} \end{cases}$$

- We denote by $y^{\#} = R_{I^{\epsilon_k}(w_k)}(y)$.

Projected BFGS method

- Let S be an index set and R_S the matrix defined by

$$R_S = \begin{cases} \delta_{ij}, & \text{if } i \in S \text{ or } j \in S; \\ 0, & \text{if not.} \end{cases}$$

- We denote by $y^{\#} = R_{I^{\epsilon_k}(w_k)}(y)$.
- H_k corresponds to the iteration matrix of the projected BFGS method

$$H_{k+1} = R_{I^{\epsilon_k}(w_k)} H_k R_{I^{\epsilon_k}(w_k)} - R_{I^{\epsilon_k}(w_k)} \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k} R_{I^{\epsilon_k}(w_k)} + \frac{y_k^{\#} (y_k^{\#})^T}{s_k^T y_k^{\#}},$$

Projected BFGS method

- Let S be an index set and R_S the matrix defined by

$$R_S = \begin{cases} \delta_{ij}, & \text{if } i \in S \text{ or } j \in S; \\ 0, & \text{if not.} \end{cases}$$

- We denote by $y^{\#} = R_{I^{\epsilon_k}(w_k)}(y)$.
- H_k corresponds to the iteration matrix of the projected BFGS method

$$H_{k+1} = R_{I^{\epsilon_k}(w_k)} H_k R_{I^{\epsilon_k}(w_k)} - R_{I^{\epsilon_k}(w_k)} \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k} R_{I^{\epsilon_k}(w_k)} + \frac{y_k^{\#}(y_k^{\#})^T}{s_k^T y_k^{\#}},$$

- We consider the recursive update of the inverse Kelly 1999:

$$B_{k+1} = \left(I - \frac{s_k^{\#}(y_k^{\#})^T}{(y_k^{\#})^T s_k^{\#}} \right) R_I B_k R_I \left(I - \frac{y_k^{\#}(s_k^{\#})^T}{(y_k^{\#})^T s_k^{\#}} \right) + \frac{s_k^{\#}(s_k^{\#})^T}{(y_k^{\#})^T s_k^{\#}},$$

Implementation details

- We solve in parallel N independent data assimilation optimality systems and N independent adjoint systems. The information is then integrated via the gradient formula
- The projected BFGS is used for the update of the placement vectors w and σ. The method is initialized with the identity matrix.
- Projected line-search rule with backtracking of the form $\frac{1}{2^k \|\nabla f(w_0)\|}, \; k = 0, 1, \ldots$
- The nonlinearity considered as example is $g(y) = \frac{y}{\sqrt{y^2+\varepsilon^2}}$

Experiment 1

- Placement is allowed in any grid point
- Observations are taken in every time step
- Goal: verify the descent of the cost functional along the iterations
- Goal: observe how the solution structure changes with respect to γ

Experiment 1

- Placement is allowed in any grid point
- Observations are taken in every time step
- Goal: verify the descent of the cost functional along the iterations
- Goal: observe how the solution structure changes with respect to γ

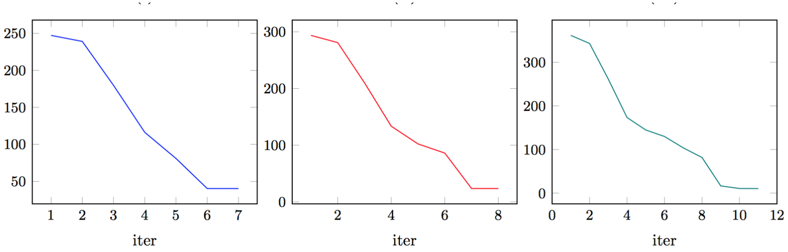


Figure: Descent of the cost function

Experiment 1

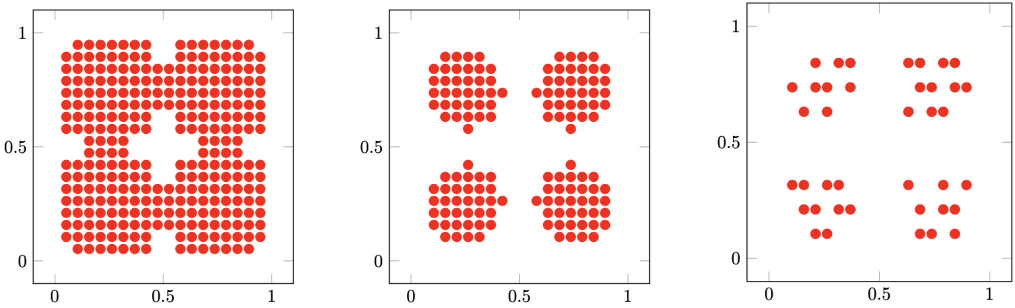
ModeMat

- Placement is allowed in any grid point
- Observations are taken in every time step
- Goal: verify the descent of the cost functional along the iterations
- Goal: observe how the solution structure changes with respect to γ

β_w	# zeros in w	# ones in w
1×10^{-5}	0	400
0.0001	88	312
0.0005	116	284
0.0020	176	224
0.0050	228	172
0.0072	268	132
0.0073	302	98
0.0074	334	66
0.0079	361	39
0.0080	400	0

Experiment 1

- Placement is allowed in any grid point
- Observations are taken in every time step
- Goal: verify the descent of the cost functional along the iterations
- Goal: observe how the solution structure changes with respect to γ



Experiment 2

- Placement is allowed only at 8 candidate locations

$$
\begin{aligned}
x_1 &= (0.2, 0.2) & x_4 &= (0.8, 0.0) & x_7 &= (0.4, 0.9) \\
x_2 &= (0.5, 0.4) & x_5 &= (0.8, 1.0) & x_8 &= (0.3, 0.8) \\
x_3 &= (0.7, 0.3) & x_6 &= (0.8, 0.6) & &
\end{aligned}
$$

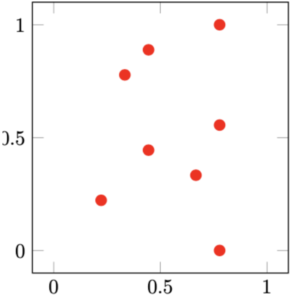
- Observations are taken in every time step
- Goal: observe how the solution structure changes with respect to the penalization parameters

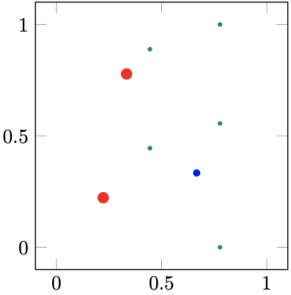
Experiment 2

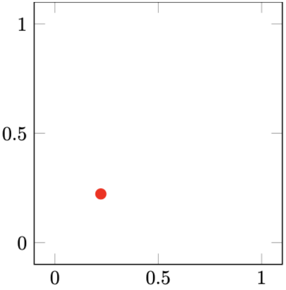
- Placement is allowed only at 8 candidate locations

$$x_1 = (0.2, 0.2) \qquad x_4 = (0.8, 0.0) \qquad x_7 = (0.4, 0.9)$$
$$x_2 = (0.5, 0.4) \qquad x_5 = (0.8, 1.0) \qquad x_8 = (0.3, 0.8)$$
$$x_3 = (0.7, 0.3) \qquad x_6 = (0.8, 0.6)$$

- Observations are taken in every time step
- Goal: observe how the solution structure changes with respect to the penalization parameters







Outline

Summary

- 4D data assimilation problems in infinite dimensions: control spaces, type of observations, nonlinear dynamics;
- Bilevel learning (data-driven) approaches for estimating parameters in PDE-constrained optimization problems;
- Optimal placement of observations/sensors in inverse problems;
- Sparse solutions for PDE-constrained optimization problems;
- Singular control problems with measures as controls;
- Efficient numerical strategies for solving bilevel PDE-constrained optimization problems.

IFIP TC7 Conference

August 31 - September 4, 2020

Save the date!