p-norm regularization in variational data assimilation

Antoine Bernigaud[a], Serge Gratton[a], Flavia Lenti[b], Ehouarn Simon[a], and Oumaima Sohab[c]

[a]Université de Toulouse, INP, IRIT, Toulouse, France
[b]CLC space, Alsbach-Hähnlein, Germany
[c]Lehigh University, Bethlehem, Pennsylvania, United States

Outline

Motivation and objective

Theoretical set up

Studied problem

Results

Conclusion

Outline

Motivation and objective

Theoretical set up

Studied problem

Results

Conclusion

Motivation and objective

- So far only the l_1-norm, the l_2-norm or a mixed of both ([Freitag, Nichols e Budd 2013]) have been used as penalty function for the formulation of the $4DVAR$ objective function (to take into account the sparsity of the variables encountered in Data Assimilation).
- However the $l_2 - norm$ tends to "oversmooth" the solution and the $l_1 - norm$ tends to "oversparcify" it. Can we make a compromise between the 2 ?
- Moreover, while a l_2-norm penalization in the $4DVAR$ objective function can be statistically interpreted as a Gaussian distribution of the errors, data can better follow a generalized Gaussian distribution ([Asadi, Scott e Clausi 2019]).

Objective

Show the benefits of using a l_p-norm with $1 < p < 2$ on a data assimilation example.

Outline

Motivation and objective

Theoretical set up

Studied problem

Results

Conclusion

Formulation of the problem

The objective function to minimize take the following form :

$$\arg\min_{\xi} \frac{1}{2} \parallel A\xi - b \parallel_2^2 + \frac{\lambda}{p} \parallel \xi \parallel_p^p = \arg\min_{\xi} \Omega_p(\xi, b, \lambda) \tag{1}$$

Where we set

$$A = \begin{pmatrix} R^{-\frac{1}{2}} \hat{H} \\ B^{-\frac{1}{2}} \end{pmatrix} \Phi^{-1}; \quad b = \begin{pmatrix} R^{-\frac{1}{2}} y \\ B^{-\frac{1}{2}} x_b^0 \end{pmatrix}, \tag{2}$$

with

- R and B the covariance matrix of the observations and the background respectively
- $\hat{H}$ is the linearization of the observation operator
- x_b^0 is the background vector and y the observations vector

Sparcity will be expected on the derivative of the variable, hence $\xi = \Phi u$ with Φ an operator of (numerical) derivation :

$$\Phi = \begin{bmatrix} 1 & & & 0 \\ -1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & -1 & 1 \end{bmatrix} \tag{3}$$

Minimization algorithm

To minimize (1) we used the algorithm proposed in [Bonesky et al. 2007]. Indeed, we obtained faster convergence with this algorithm rather than with a plain gradient descent. One iteration is done in two steps :

$$\begin{cases} \xi_k^* = j_p(\xi_{k-1}) - \mu_k \nabla \Omega_p(\xi_{k-1}, b, \lambda). \\ \xi_k = j_q(\xi_k^*). \end{cases} \tag{4}$$

where

- $j_p : \mathbb{R}^n \to \mathbb{R}^n$ is the so-called "duality map", which is the derivative of $\xi \to \frac{\lambda}{p} \parallel \xi \parallel_p^p$ and whose expression is here reduced to
$$j_p(\xi)_i = |\xi_i|^{p-1} \text{sign}(\xi_i) \tag{5}$$
- q is such that $\frac{1}{p} + \frac{1}{q} = 1$

Choice of the regularization parameter λ

This choice is based on Morozov's discrepancy principle [Anzengruber e Ramlau 2009-2013]. Let $\tau \geq 1$ and $b_{no-noise} \in rg(\mathbf{A})$ be the data without noise. For $\delta > 0$ and with $\| b - b_{no-noise} \| \leq \delta$, we chose the regularization parameter $\lambda > 0$ if there exists ξ_λ^δ such that:

$$\left\{ \begin{array}{c} \xi_\lambda^\delta = \arg\min_\xi \Omega_p(\xi, b, \lambda) \\ \| \mathbf{A}\xi_\lambda^\delta - b \|_2^2 \leq \tau\delta. \end{array} \right. \tag{6}$$

In practice we compute λ by backtracking : since $\lambda \to \| A\xi_\lambda^\delta - b \|_2^2$ is increasing, we start with a λ_0 and update $\lambda_{k+1} = 0.8\lambda_k$ as long as (6) is not fulfilled.

Outline

Motivation and objective

Theoretical set up

Studied problem

Results

Conclusion

Studied problem

The experiment consist in retrieving the initial condition, given noisy observations and a background information, of the following 1D-advection problem :

$$\left\{ \begin{array}{c} \partial_t u(s,t) + c\partial_s u(s,t) = 0 \\ u(s,t_0) = u_0(s) \\ u(0,t) = u(L,t) = 0 \end{array} \right. \tag{7}$$

with c = L = 1

Perfect VS Imperfect

We discretize the equation 7 using the Lax-Wendroff scheme ([Lax e Wendroff 1960]). By writing Δt and Δx the time and space steps of the discretization, this numerical model give rises to implicit diffusion when $\mu = c\frac{\Delta t}{\Delta x} < 1$ and will modify the sparcity of the variables as time goes by, as shown below

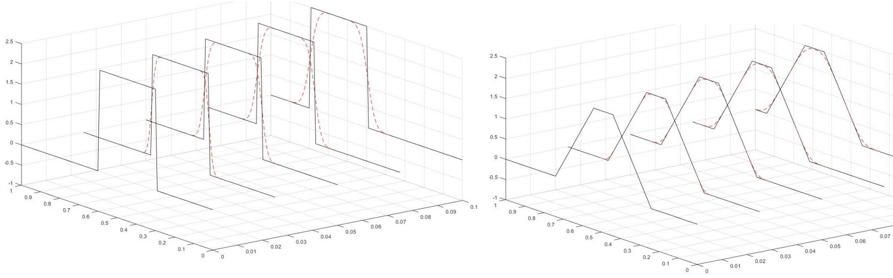
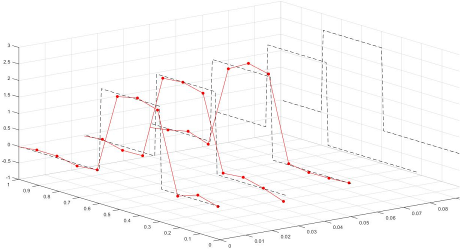
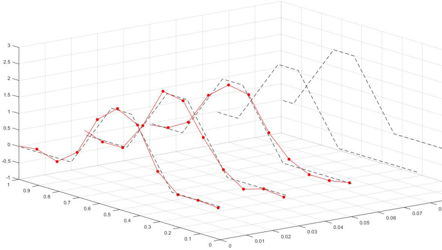


Figure 1:

Sparse VS Almost Sparse

We also consider two types of initial condition : a sparse and an "almost sparse" one. Below is example of noisy measurements taken for each case. Note that the sparcity concerns the derivative of the variables.





Summary of the problem

We then investigate 4 cases : with a sparse initial condition and with and almost sparse one, with $\mu = 1$ (called "perfect scenario") and with $\mu = 0.5$ ("imperfect scenario").
To palliate the randomness that takes place in the experiments (when creating the background and the observations data), we perform a minimization of the objective function 20 times for each case and we report the mean of the RMSE and MAE.

Outline

Motivation and objective

Theoretical set up

Studied problem

Results

Conclusion

Sparse scenario

B	R			background	$4DVAR$	$4DVAR,1$	$4DVAR,1.2$	$4DVAR,1.5$	$4DVAR,2$
$0.1I$	$0.1I$	Sparse	RMSE	0.2886	0.2427	<u>0.1153</u>	0.1667	0.1834	0.2063
			RMAE	0.4231	0.3530	<u>0.1672</u>	0.1980	0.2050	0.2343
		Al. sparse	RMSE	0.2809	0.2388	<u>0.1191</u>	0.1370	0.1438	0.1568
			RMAE	0.3727	0.3119	<u>0.1496</u>	0.1771	0.1738	0.1845
$0.01I$	$0.1I$	Sparse	RMSE	0.0901	0.0878	0.0748	<u>0.0636</u>	0.0804	0.1017
			RMAE	0.1331	0.1296	0.1089	<u>0.0636</u>	0.0828	0.1124
		Al. sparse	RMSE	0.0875	0.0853	0.0742	<u>0.0615</u>	0.0668	0.0766
			RMAE	0.1174	0.1144	0.0983	0.0717	<u>0.0694</u>	0.0819
$0.1I$	$0.01I$	Sparse	RMSE	0.2869	0.2014	0.2210	<u>0.1141</u>	0.1310	0.1475
			RMAE	0.4254	0.2619	0.3621	<u>0.1028</u>	0.1147	0.1393
		Al. sparse	RMSE	0.2715	0.1904	0.2148	<u>0.0852</u>	0.0884	0.0973
			RMAE	0.3568	0.2233	0.3227	0.1044	<u>0.0956</u>	0.1059

Table 1: Perfect model scenario : RMSE and MAE related to the sparse and almost sparse cases experiments. The best result for each row is underlined.

When uncertainty increase, the l_p-norms ($p = 1.2$ and $p = 1.5$ give better results.

Sparse scenario

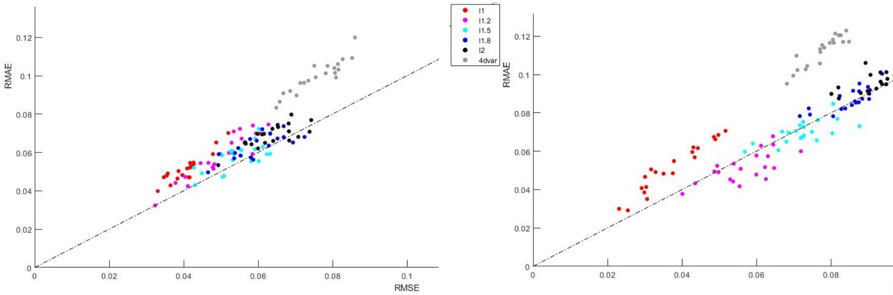


Figure 2: Distribution of the RMSE and MAE of 20 experiments for the perfect scenario and $R = 0.1$; $B = 0.1$. On the left : the almost sparse case ; on the right : the sparse case. The almost sparse case "gathers" the points to the benefits of the l_p-norm.

Sparse perfect scenario

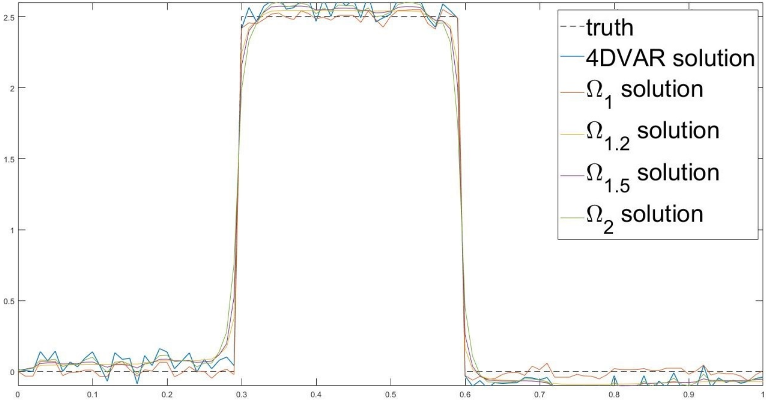


Figure 3: The different solutions to the minimization of the different objective function. Oscillations increase as p increase.

Sparse imperfect scenario

	$t = 0$		$t = 0.025$		$t = 0.05$	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
imperfect	0	0	0.1526	0.0820	0.1841	0.1175
background	0.2943	0.4313	0.1947	0.2393	0.2112	0.2430
$4DVAR$	0.2746	0.3981	0.1738	0.1991	<u>0.1930</u>	0.2035
$4DVAR$,1	0.2431	0.3486	<u>0.1734</u>	0.1973	0.1931	<u>0.2032</u>
$4DVAR$,1.2	<u>0.1195</u>	<u>0.1331</u>	0.1876	<u>0.1808</u>	0.2119	0.2066
$4DVAR$,1.5	0.1454	0.1617	0.1946	0.1953	0.2154	0.2161
$4DVAR$,2	0.1754	0.2003	0.2037	0.2120	0.2202	0.2267

Imperfect model scenario: RMSE and MAE related to the sparse case experiment. Underlined are the best results for each column. While the $l_{1.2}-$penalty lead to the better results for $t = 0$, the model error tends to reduce the gap between the different penalization as time goes by.

Almost sparse imperfect scenario

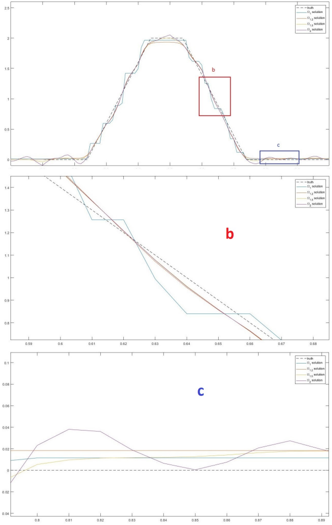


Figure 4: Imperfect model scenario: initial true state (black dotted line) and initial state obtained minimizing $\Omega_{\mathbf{1}}$ (blue line), $\Omega_{\mathbf{1.2}}$ (red line), $\Omega_{\mathbf{1.5}}$ (yellow line) and $\Omega_{\mathbf{2}}$ (violet line). b (second picture) shows the $l_{\mathbf{1}}$-norm produces a "staircase effect" on the reconstruction, but c (third picture) shows using a $p-$norm with p close to 2 leads to oscillations.

Outline

Motivation and objective

Theoretical set up

Studied problem

Results

Conclusion

Conclusion

- The l_p-norm ($1 < p < 2$) is able to better reconstruct "almost" sparcity
  $\rightarrow$ can be useful for
  - ocean ice models (e.g. derivative of the ice concentration in the marginal ice zone)
  - atmospheric models (e.g. cloud coverage)
- How to tune the parameter λ in (1) ? We used the Morozov discrepancy principle for the experiments (see Annex 2), but it requires the solving of several optimization problems. That may be impracticable in actual data assimilation problem.
- What algorithm is the most suited to minimize the objective function ?

Bibliography

ANZENGRUBER, S.; RAMLAU, R. Morozov's discrepancy principle for tikhonov-type functionals with non-linear operators. *Johann Radon Institue for Computational and Applied Mathematics Austrian Academe of Sciences*, RICAM-Report, 2009–2013.

ASADI, N.; SCOTT, K. A.; CLAUSI, D. A. Data fusion and data assimilation of ice thickness observations using a regularisation framework. *Tellus A: Dynamic Meteorology and Oceanography*, 2019.

BONESKY, T. et al. Minimization of tikhonov functional in banach spaces. Abstract and applied analysis, v. 2008, p. 1–18, 2007.

FREITAG, M.; NICHOLS, N.; BUDD, C. Resolution of sharp fronts in the presence of model error in variational data assimilation. *Quartely journal of the royal meteorological society*, v. 139, p. 742–757, 2013.

LAX, P.; WENDROFF, B. Systems of conservations laws. Commun. Pure Appl. Math., v. 13, p. 217–237, 1960.

Acknowledgements

This study has been funded by the AVENUE project of the RTRA STAE Foundation. Ehouarn Simon was partially supported by the French national program LEFE (Les Enveloppes Fluides et Environnement).