L_p-norm regularization with $1 < p < 2$ in variational data assimilation

Antoine BERNIGAUD, Serge GRATTON, Ehouarn SIMON

Toulouse university, INP, IRIT, France

Introduction

The L_1 and L_2 norms have been successful as regularization terms in data assimilation (Freitag *et al.*, 2010). The first one promotes a sparse solution while the second one promotes a smoother solution. The solution may however possess a structure "in between" that we call "quasi-sparse". The L_p-norm with $1 < p < 2$ aims at making a compromise between these 2 norms. Moreover, the L_2 and L_1 norms introduce oscillations in the solution, and it has been shown that considering the L_p-norm with $1 < p < 2$ can mitigate these oscillations (Schuster *et al.*, 2012). Finally, the use of the L_p-norm is also motivated by the statistical distribution of physical variables, when it follows a generalized Gaussian distribution instead of the Laplace distribution or the more classical Normal distribution.

The penalized 4DVar

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}^b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\mathbf{y} - \hat{\mathcal{H}}(\mathbf{x})\|^2_{\mathbf{R}^{-1}} + \frac{\lambda}{p}\|\mathbf{\Phi}(\mathbf{x})\|_p^p \qquad (1)$$

The regularization can be in any basis of interest

The operator $\mathbf{\Phi}$ stands for the projection of $\mathbf{x}$ in the basis where sparsity is expected (a Fourier, derivative, wavelet basis...)

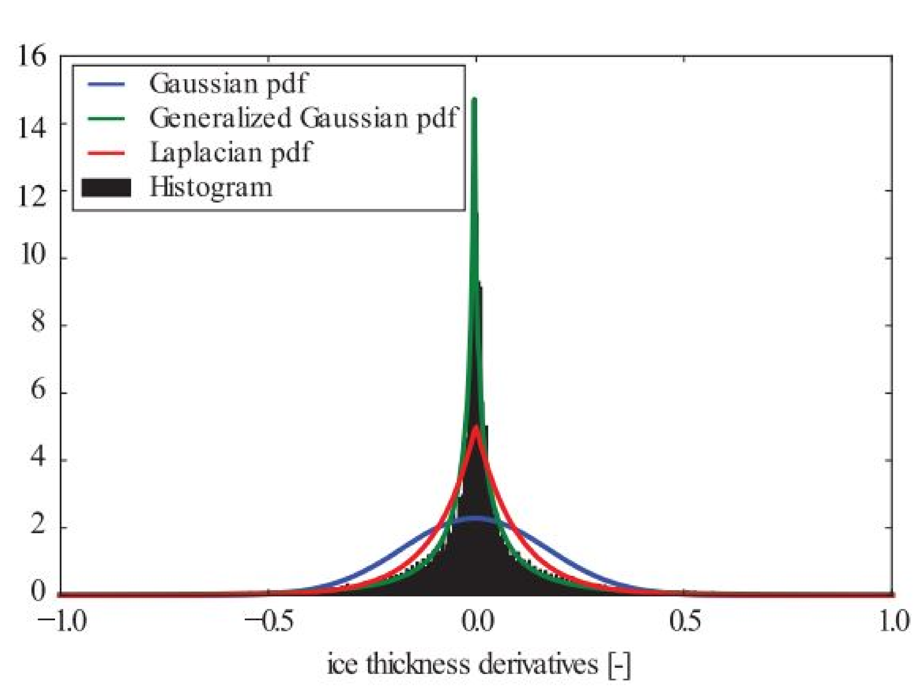
How to choose λ and p?

- If the regularization is motivated by the statistical distribution of the variables, these parameters can be derived from the modelization of the problem.
- Otherwise p depends on the expected structure of the solution. The more sparse the solution is, the closer to 1 p should be chosen.
- λ can be estimated with the L-curve method or the Morozov discrepancy principle.

Statistical benefits of the regularization

The distribution of the variable may follow a generalized Gaussian distribution: cases of

- the underwater acoustic noise (Banerjee *et al.*, 2013).
- the wind velocity gradient (Stengel *et al.*, 2019)
- the derivative of the Beaufort sea ice concentration (Asadi *et al.*, 2019).



The PDF of the generalized Gaussian distribution

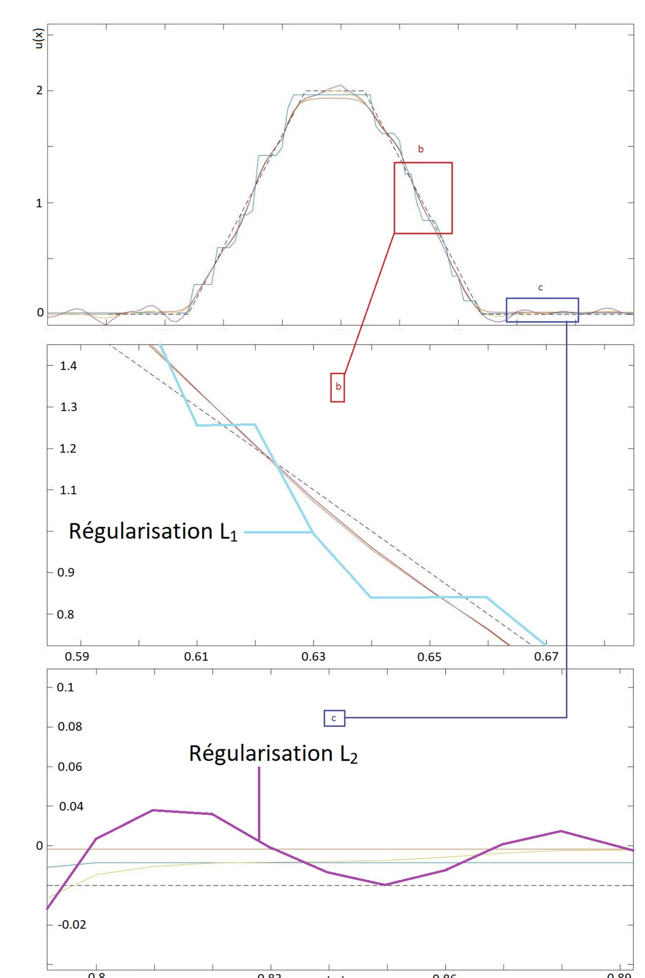
$$f(x;\alpha,p) = \frac{p}{2\alpha\Gamma(1/p)}\, e^{-\left(\frac{|x-\mu|}{\alpha}\right)^p}$$

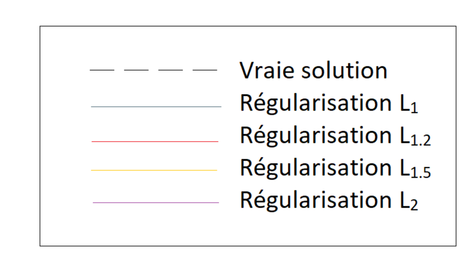
with μ position parameter, α scale parameter, p shape parameter.

Assessment of the numerical benefits on data assimilation experiments

We show the benefits of this regularization on two data assimilation setups: a 1D toy model based on the advection equation, which allowed to study the effect of the penalization on an easily tuned experiment, and a more realistic 2D experience based on the shallow water equations, which allowed to both put to the test the behavior of the regularization and highlight the algorithms proposed to minimize (1) for a higher dimension problem. The results can be found in [1] and [2].

A 1D advection data assimilation experiment





Zoom on b : "Staircase effect" introduced by the L_1-norm.

Zoom on c : Oscillations introduced by the L_2-norm.

A 2D shallow water data assimilation experiment

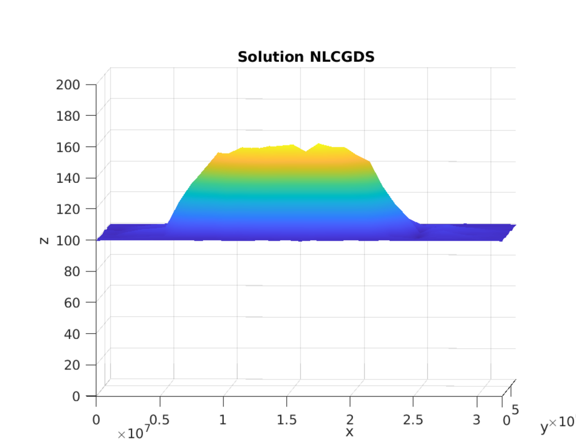


Figure 1.

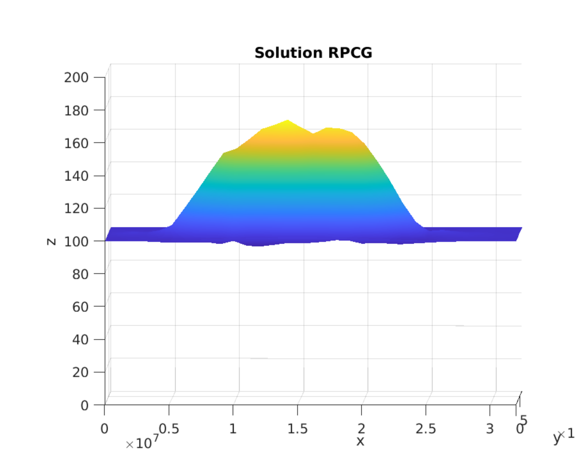


Figure 2.

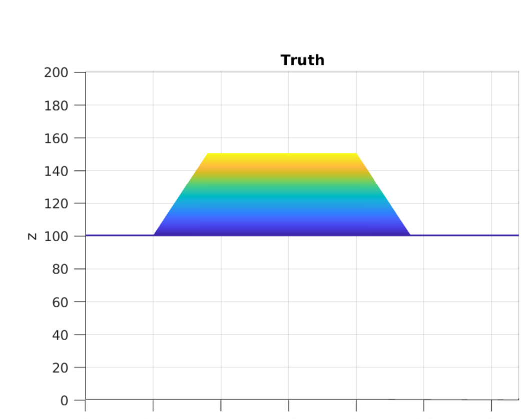


Figure 3.

- Figure 1 : Analysis obtained by minimizing the regularized 4DVar with a dual-space non linear conjugate gradient.
- Figure 2 : Analysis obtained by minimizing the 4DVar without L_p-norm regularization with preconditioned conjugate gradient.
- Figure 3 : Truth.

Quasi-sparse signals in practical applications

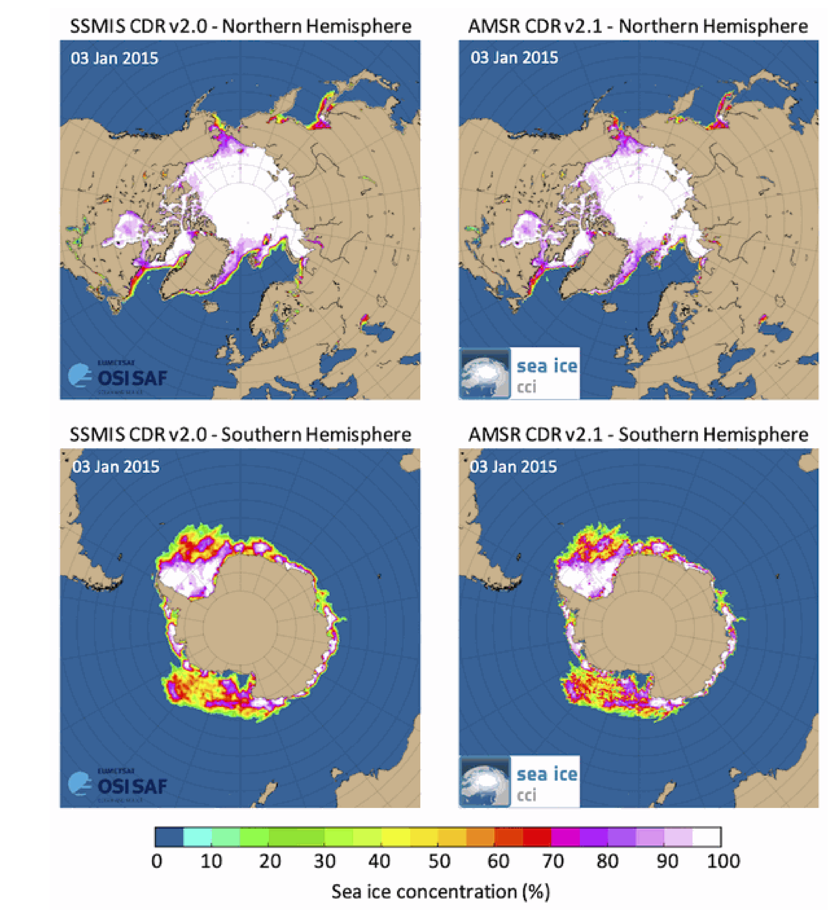


Figure 4. Sea ice concentration (data collected by EUMETSAT)

other examples of quasi-sparse signals: Plankton concentration (A. Samuelsen *et al.*, *Hurricane-driven alteration in plankton community size structure in the Gulf of Mexico: A modeling study*), sargassum concentration, meteorological fronts (Freitag *et al.*, 2013)...

How to minimize the penalized 4DVar?

A plain gradient descent would work but is slow. The right mathematical framework to minimize a functional penalized by an L_p-norm with $1 < p < 2$ is a Banach space (Schuster *et al.*, 2012).

How to do a gradient descent in a Banach space?

Suppose $x \in L_p$. The following equation is not well defined:

$$x_{k+1} = \underbrace{x_k}_{\in X} + \alpha_k \underbrace{-f'_k}_{\in X^*}. \qquad (2)$$

We need to transport the direction in the topological dual space (L_q, with $\frac{1}{p}+\frac{1}{q}=1$):

$$p_0 = -f'_0, \qquad (3)$$
$$x^*_{k+1} = x^*_k + \alpha_k p_k, \qquad (4)$$
$$x_{k+1} = J_q(x^*_{k+1}), \qquad (5)$$
$$p_{k+1} = -f'_{k+1}. \qquad (6)$$

Or to transport the iterate in the primal space:

$$p_0 = -J_q(f'_0) \qquad (7)$$
$$x_{k+1} = x_k + \alpha_k p_k, \qquad (8)$$
$$p_{k+1} = J_q(-f'_{k+1}). \qquad (9)$$

References

[1] A. Bernigaud, S. Gratton, F. Lenti, E. Simon, and O. Sohab. Lp-norm regularization in variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 147:2067–2081, 2021.

[2] A. Bernigaud, S. Gratton, and E. Simon. A non-linear conjugate gradient in dual space for L_p-norm regularized non-linear least squares with application in data assimilation. *To be published.*

antoine.bernigaud@nersc.no