L^p-norm regularization in variational data assimilation

Antoine Bernigaud[‡,•], Serge Gratton[‡], Ehouarn Simon[‡]
‡Université de Toulouse, INP, IRIT, Toulouse, France • now at NERSC, Bergen, Norway

Motivations

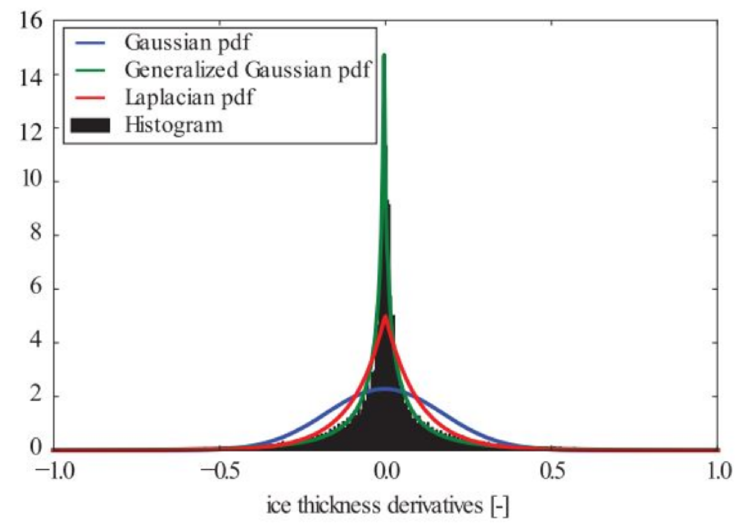
- The L^1 and L^2 norms have been successful as regularization terms in data assimilation (Freitag et al., 2010, 2013), but can introduce oscillations in the solution (Schuster et al., 2012).
- The L^p-norm with $1 < p < 2$ aims at making a compromise between these 2 norms ("quasi-sparse" solution) and can mitigate the occurence of spurious oscillations (Schuster et al., 2012).
- The use of the L^p-norm is also motivated by the statistical distribution of physical variables following a generalized Gaussian distribution.

Generalized Gaussian distribution

- Probability density function:

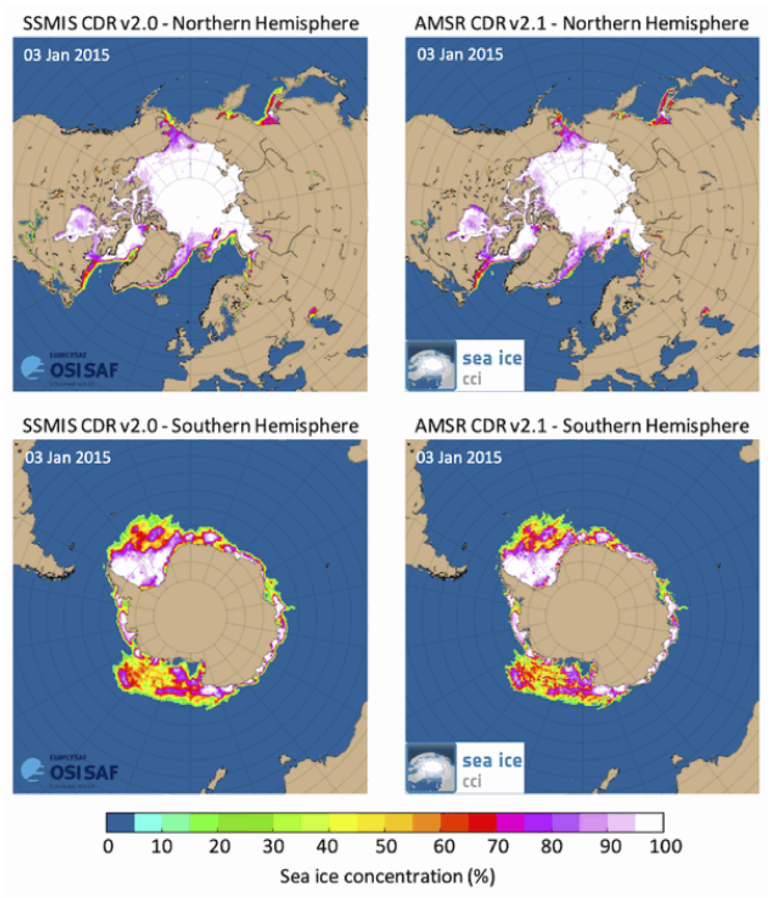
$$f(x; \alpha, \mu, p) = \frac{p}{2\alpha\Gamma(1/p)} e^{-\left(\frac{|x-\mu|}{\alpha}\right)^p},$$

with $\mu \in \mathbb{R}$ position parameter, $\alpha > 0$ scale parameter and $p > 0$ shape parameter.
- Special case : Laplace distribution ($p = 1$) and Normal distribution ($p = 2$).
- p close to 1: sparsity.
- Example: derivative of the Beaufort sea ice concentrations (Asadi et al., 2019)

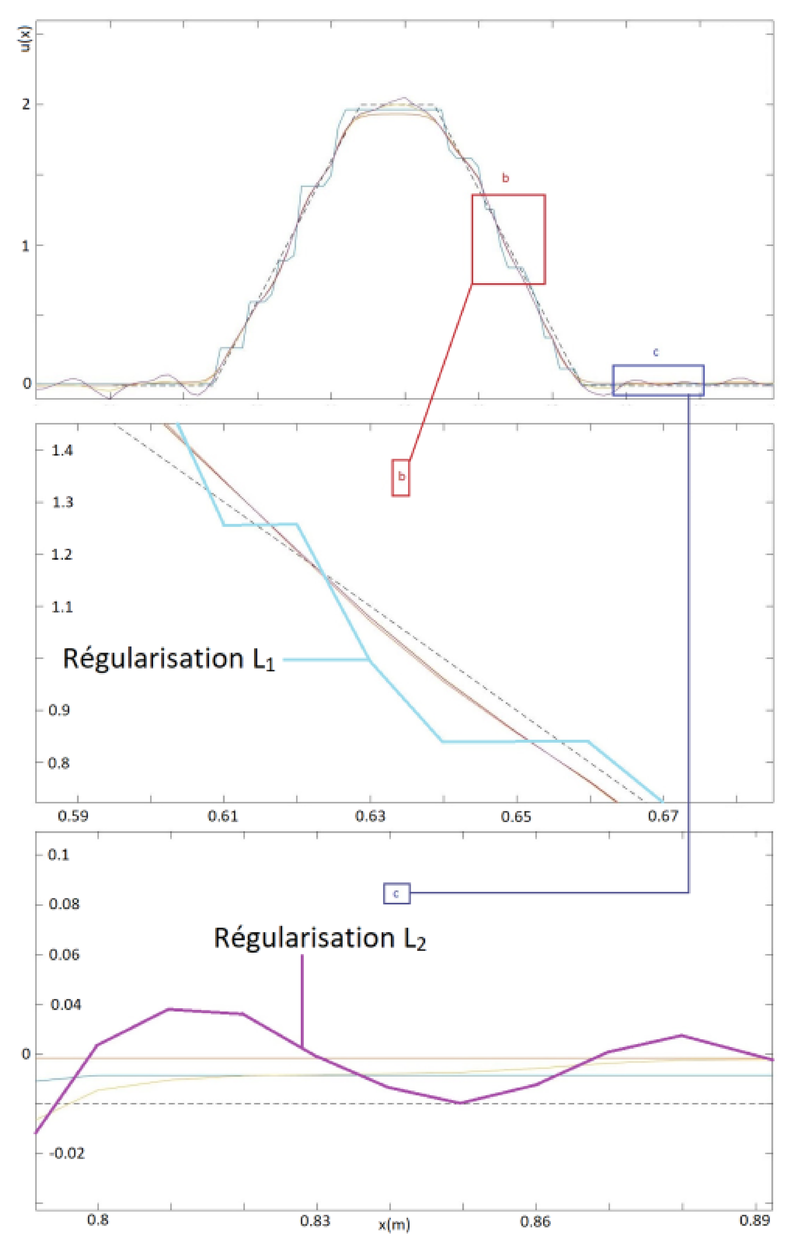


"Almost sparse" variables

- In between sparse and smooth variables.
- Example: sea ice concentration (EUMETSAT).



- Variational data assimilation in a 1D linear advection model (Bernigaud et al., 2021).



- Staircase effect (L^1-norm) and oscillations (L^2-norm)
- Better results with $p = 1.2$ but slow decrease of the cost function.

Penalization of the 4D-var cost function

$$\min_{\mathbf{x}_0 \in \mathbb{R}^n} f(\mathbf{x}_0) = \frac{1}{2}\|\mathbf{y} - \mathcal{H}(\mathcal{M}(\mathbf{x}_0))\|^2_{\mathbf{R}^{-1}} + \frac{1}{2}\|\mathbf{x}_b - \mathbf{x}_0\|^2_{\mathbf{B}^{-1}} + \frac{\lambda}{p}\|\Phi\mathbf{x}_0\|^p_p$$

with
- $\mathbf{x}_0 \in \mathbb{R}^n$ the initial condition, $\mathbf{x}_b \in \mathbb{R}^n$ the background state vector and $\mathbf{y} \in \mathbb{R}^m$ the observations;
- $\mathcal{H}$ the observation operator, $\mathcal{M}$ the model;
- $\lambda > 0$ the weight of the regularization, $1 < p < 2$, and Φ a linear operator (projection on the Fourier or wavelet basis, derivative...).

Minimization in Banach spaces $(\mathbb{R}^n, \|\|_p)$

- First-order descent algorithm:

$$\mathbf{x}_{k+1} = \underbrace{\mathbf{x}_k}_{\in \mathbb{R}^n} - \alpha_k \underbrace{f'(\mathbf{x}_k)}_{\in \mathcal{L}_c(\mathbb{R}^n, \mathbb{R})}.$$

- $f'(\mathbf{x}_k)$ can be identified with an element of $(\mathbb{R}^n, \|\|_q)$, $\frac{1}{p} + \frac{1}{q} = 1$.
  - Well defined iteration if $p = q = 2$ (Hilbert space).
  - Numerical experiments in 1D: huge number of iterations if $p \neq q$ (Bernigaud et al., 2021).
- Transport the direction in $(\mathbb{R}^n, \|\|_q)$ (Schuster et al., 2012):

$$\begin{cases} \mathbf{x}^*_{k+1} = \mathbf{x}^*_k - \alpha_k f'(j_q(\mathbf{x}^*_k)) \\ \mathbf{x}_{k+1} = j_q(\mathbf{x}^*_{k+1}) \end{cases}$$

with $\mathbf{x}^*_k = j_p(\mathbf{x}_k)$, and the duality map $j_p : (\mathbb{R}^n, \|\|_p) \to (\mathbb{R}^n, \|\|_q)$ defined by

$$\forall i = 1 \cdots n, \quad [j_p(\mathbf{x})]_i = \text{sign}(\mathbf{x}_i)|\mathbf{x}_i|^{p-1}.$$

  - Numerical experiments in 1D : significant reduction in the number of iterations (Bernigaud et al., 2021).

A nonlinear conjugate gradient in dual space

- Solving a nonlinear least square problem

$$f(\mathbf{x}) = \frac{1}{2}\|A(\mathbf{x}) - \mathbf{b}\|^2_2 + \frac{\lambda}{p}\|\Phi\mathbf{x}\|^p_p,$$

with A nonlinear.
- Global convergence property:
  - Conditions on the length step α_k.
- Faster decay of f over iterations:
  - Choice of a direction $\mathbf{p}_{k+1} = -f'(j_q(\mathbf{x}^*_{k+1})) + \beta_k\mathbf{p}_k$, with β_k to be defined.

Algorithm NLCGDS (Bernigaud et al., 2023)

- Wolfe-like conditions on α_k in dual space:

$$f_{k+1} = (f \circ J_q)(x^*_k + \alpha_k p_k) \le f_k + c_1\alpha_k\langle\nabla f_k, J'_q(H_k(x^*_k))p_k\rangle \quad (13)$$

$$\langle\nabla f_{k+1}, J'_q(H_k(x^*_k))p_k\rangle \ge c_2\langle\nabla f_k, J'_q(H_k(x^*_k))p_k\rangle. \quad (14)$$

- Full algorithm:

Algorithm 2 NLCGDS with step search in the dual space

Choose c, c_0 in $(0;1)$, x_0, p in $(1;2)$ and $n^\beta_{itermax}$, $K \in \mathbb{N}^*$
$x^*_0 \leftarrow J_p(x_0)$
$p_0 \leftarrow -\nabla f_0$
for $k = 0..N-1$ **do**
  Computation of α_k that satisfies (13) and (14)
  $x^*_{k+1} \leftarrow x^*_k + \alpha_k p_k$
  if k is a multiple of K **then**
    $\beta_k = 0$
  else
    $y^*_k = \nabla(f \circ J_q)(x^*_{k+1}) - \nabla(f \circ J_q)(x^*_k)$
    $\beta_k \leftarrow \frac{\langle\nabla f_{k+1}, y^*_k\rangle}{\langle p_k, y^*_k\rangle}$
    $l \leftarrow 0$
    while $\langle(f \circ J_q)'(x^*_k), -\nabla f_{k+1} + \beta_k p_k\rangle > 0$ and $l < n^\beta_{\text{itermax}}$ **do**
      $\beta_k \leftarrow c\beta_k$
      $l \leftarrow l + 1$
    end while
    if $l = n^\beta_{itermax}$ **then**
      $\beta_k \leftarrow 0$
    end if
    $p_{k+1} \leftarrow -\nabla f_{k+1} + \beta_k p_k$
  end if
end for
Return x_N

Bibliography

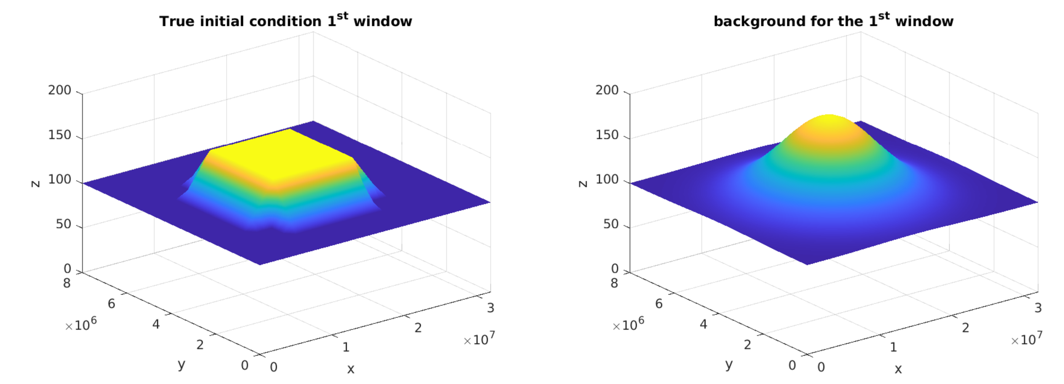
- Asadi N., Scott K.A., Clausi D.A.: Data fusion and data assimilation of ice thickness observations using regularisation framework, *Tellus A*, 71(1), 2019.
- Bernigaud A., Gratton S., Lenti F., Simon E., Sohab O.: Lp-norm regularization in variation data assimilation, *Q.J.R. Meteorol. Soc.*, 142, 2021.
- Bernigaud A., Gratton S., Simon E.: A nonlinear conjugate gradient in dual space for Lp-norm regularized nonlinear least squares with application in data assimilation, *Numer. Algor.*, 2023.
- Freitag M., Nichols N., Budd C.: L1-regularization for ill-posed problems in variational data assimilation, *Proc. Appl. Math. Mech.*, 10, 2010.
- Freitag M., Nichols N., Budd C.: Resolution of sharp fronts in the presence of model error in variational data assimilation, *Q.J.R. Meteorol. Soc.*, 139, 2013.
- Schuster T., Kaltenbacher B., Hofmann B., Kazimierski K.S.: Regularization methods in Banach spaces, *Radon series on Computational and Applied Mathematics*, 2012.

Numerical experiments

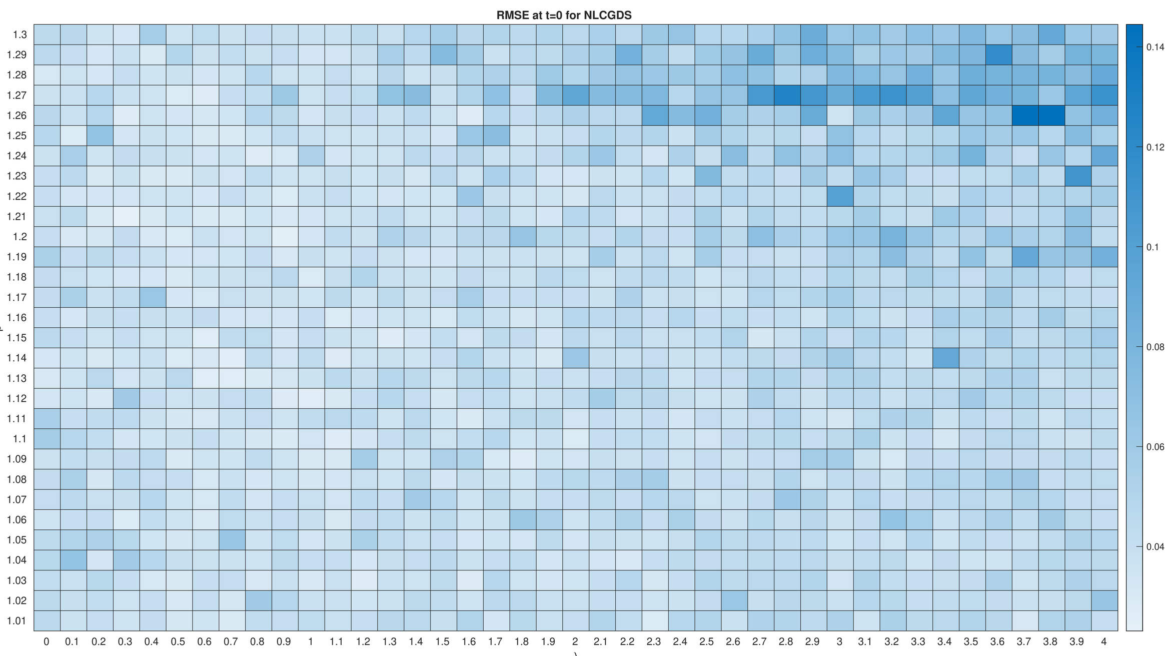
- A 2D shallow water model:

$$\begin{cases} \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv + g\frac{\partial h}{\partial x} = 0, \\ \frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} - fu + g\frac{\partial h}{\partial y} = 0, \\ \frac{\partial h}{\partial t} + u\frac{\partial h}{\partial x} + v\frac{\partial h}{\partial y} + h(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) = 0 \end{cases}$$

- A 6-hours Assimilation window.
- Observations: h, every 10 time step and 4% of the spatial domain (random selection).
- Diffusion-based modeling of **B** and diagonal **R**.
- True initial condition and background:



Choice of λ and p: heat map of RMSE

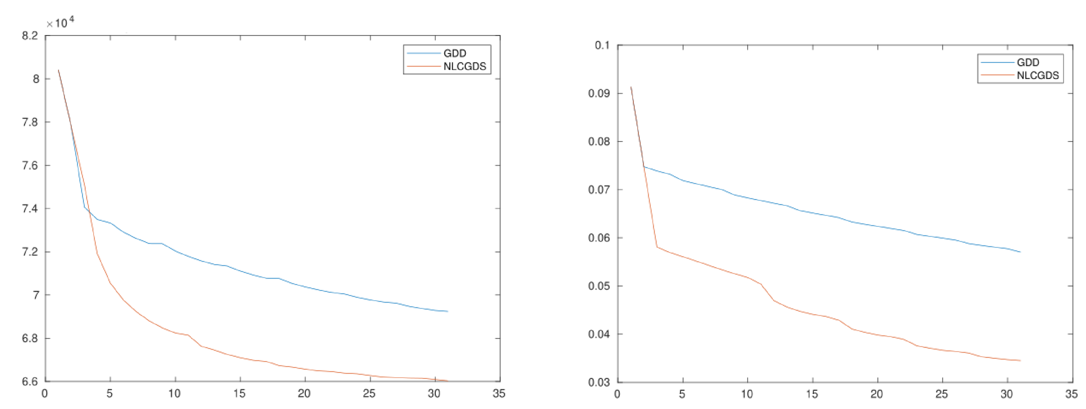


- Arbitrary choice: $p = 1.1$ et $\lambda = 1$.

Decay of the cost function and RMSE in the initial condition

- Gradient in dual space (GDD, blue) VS NLCGDS (red)

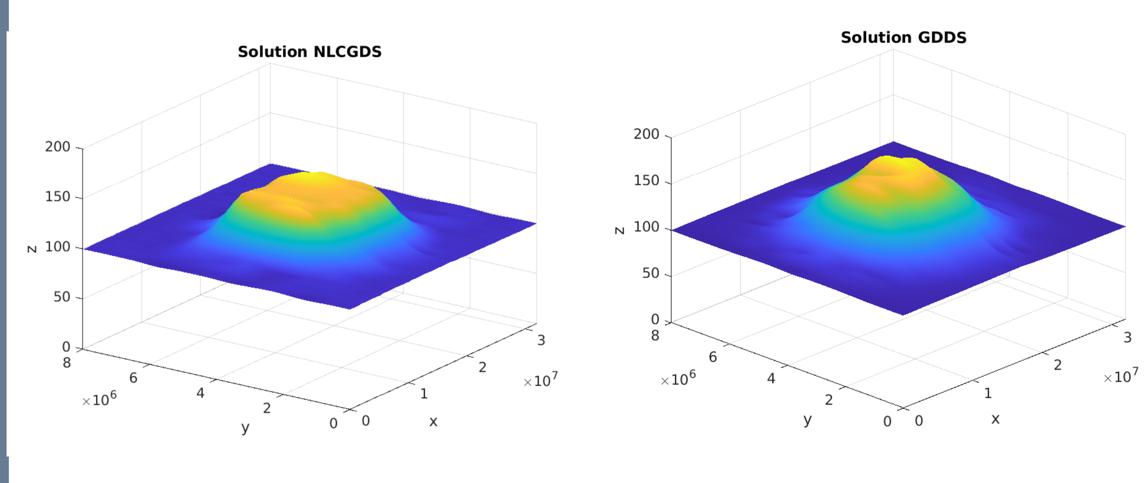
Cost function RMSE at t=0



- Fastest decay of the cost function and RMSE with NLCGDS.

Optimized initial condition (30 iterations)

- Better representation of the quasi-sparse structure with NLCGDS.



Conclusion and perspectives

- Conclusion:
  - Benefits of the L^p-norm regularization for promoting sparsity.
  - Algorithmic developments for the efficient resolution of L^p-norm regularized least square.
- Perspectives:
  - Numerical experiments in a realistic configuration.
  - Development of ensemble-based algorithms.

Acknowledgement

This study was partly supported by the French national program LEFE - MANU.

ehouarn.simon@toulouse-inp.fr