Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2017-46-AC1, 2017

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

Interactive comment on "Optimal Transport for Variational Data Assimilation" *by* Nelson Feyeux et al.

Nelson Feyeux et al.

arthur.vidard@inria.fr

Received and published: 13 November 2017

a revised version of the paper is attached as a supplement

Reply to referee 1 (RC1)

We would like to thank the referee for his/her extensive review on our paper and for giving us the opportunity to improve our paper.
We copied your commentary in italics below, we reply in normal font.

Major comments :

1) The example introduced in Fig. 1 to illustrate the potential of the method is not clear and could be improve as follows: a) You should precise the distribution name within

Printer-friendly version

Discussion paper

the paragraph: "This is illustrated in Fig. 1 which shoes two densities ρ_0 and ρ_1. The second density ρ_1 can be seen as the first one ρ_0 with position error." ;

Corrected, thank you.

b) I guess the terminology of density & distribution & probability distribution should be avoid to prevent from any confusion in DA application, and especially the probabilistic interpretation of DA (see next comments 2));

Ok. In the introduction, we replaced "density" with either "curve" (to describe ρ_0 and ρ_1) or "measure" or "mass" (in the optimal transport section). See comments 2 for the rest of the paper.

c) You should introduce the formalism for L^2 cost functions saying that the minimum of the cost function $||\rho - \rho_0||_2^2 + ||\rho - \rho_1||_2^2$ is given by $\rho_* = \frac{1}{2}(\rho_0 + \rho_1)$; while the average in the sens of the Wasserstein distance is the one of the figure, that is in between the two densities – without detailing the Wasserstein distance, as it is in the present manuscript.

Ok.

2) The work presented here is limited to the case where the state vector and observations are positive fields with finite and normalised integral – part of the state vector is assume to be a probability measure over the domain – this seems very restrictive compared with the diversity of fields usually considered in data assimilation but solution to manage this issue can be considered (especially for image data). However the restriction to being a probability measure is not my objection: the problem I see is the possible confusion between probability distribution of error (forecast and analysis error distributions) and the particular case where a field (or part of the state vector)

is a probability distribution. I think it would help the reader to insist on the difference between the classical framework of DA (with generic vector state) and this particular case, so to avoid any confusion between the particular field property (probability in compact domain in the physical space) and classical error distribution (probability in state space): while mathematically appropriate, I think the terminology of probability densities P(Omega) (section2.2.1 and definition 2.1) should be replaced by something far from "probability densities". For instance in place of "probability densities" (title section 2.2.1 & definition), you could introduce a particular class for the fields, for instance it could be called "mass-class", keeping this terminology all along the manuscript, with a remark paragraph that would precise that in optimal transport what is so-called mass-class is actually probability distribution, indicating that the terminology is introduced to prevent from confusion with state/error probability distribution.

Ok, thank you. We removed all occurrences of "densities", we replaced them by "mass functions". Following your suggestion, we included a remark in paragraph 2.2.1 Mass functions (previously "probability densities").

3) Kantorovitch potential (K-potential) plays a crucial role in the theoretical presentation as well as in the numerical solution of the minimising process, but very few is said about its computation. - How the K- potential is it computed in this study : please give the detail of the algorithm used here, the indication provided in the manuscript about the construction of the K- potential in 1D (line 1-6 p6) is not enough. Detail, at least within a paragraph, how the K- potential can be computed in 2D/3D, even if only 1D example are considered here.

Ok, we added such a paragraph detailing numerical computation of K and W2 in 1D and 2/3D, at the end of 2.2.2.

- Illustrate what is the K-potential for the particular case of two gaussian distribution

where ρ_0 (ρ_1) is a Gaussian of mean m_0 (m_1) and variance σ_0^2 (σ_1^2). If it exists, give the analytical expression for the potential in this case ?

Ok. We included such an example (Example 2.3) at the end of Section 2.2.2.

4) p12,l1-2 and l14-15: Following the author and the numerical example developed in this section, the minimising problem Eq(14) leads to two different solutions depending the choice of the dot product used along the minimising process, but no detail is given explaining why this situation occurs. This could be due to possible multiple minima of the cost function or to a non-convergence of the minimising process when using the L^2 dot product. Authors mentioned the "success of the minimisation of J''_w (l15) but without clearly indicating if the convergence was successful, or not, for the L^2 dot product. In this simple example, uniqueness of the minimum should be guaranteed, indicating that the L^2 dot-product is not able to provide a good path toward the minimum. If this is correct than the author should mention it more clearly: "In this example, the minimising process based on the L^2 scalar product fails to reach the unique minimum of the cost function as shown on ... (additional illustration)"

An additional figure (or panel in Fig.3) is needed to observe the non-convergence toward the minimum for this situation: please shows the value of the cost function J_w along the iterations of the minimising process when using the two dot-products.

I think a discussion is missing concerning existence and unicity of the J_w cost function, this should be included at the end of section 3.1.

Is it possible to replace the steepest descent by a conjugated gradient ? Do you think that this replacement could improve the convergence for the L^2 gradient ?

Ok, thank you for this helpful comment.

Regarding unicity of Jw's minimiser, we added a few sentences at the end of Section 3.1 (page 8).

Regarding conjugate gradient, we added a plot in Figure 3 comparing convergence speed of (DG#), (DG2) and a version of (DG2) using the conjugate gradient algorithm. Conjugate gradient speeds up the algorithm, but is not as fast as (DG#). See Figure 3 (page 13) and the third paragraph of 4.1 (page 12) about it.

Minor comments:

1) p1, l11: "To achieve that goal" $\rightarrow$ *"... this goal"* ***Ok***

2) p1,l17: ".. to be sought (the control vector) is .." $\rightarrow$ *".. to be sought, the control vector, is .."* ***Ok***

3) p7, l9: ω_b *is not defined in Eq(13)* ***Ok***

4) p3,l8: "Wasserstein distance is to compare" $\rightarrow$ *" Wasserstein distance to compare"* ***Ok***

5) p3,l9: "data assimilation Actual" $\rightarrow$ *"data assimilation. Actual"* ***Ok***

6) p3, l32: Observational operator is denoted by "G" in place of the more classical "H" notation. Please replace G into H along the manuscript.

Actually in our manuscript, G denote $H \circ M$, and it is a classical notation in DA. However, our phrasing was indeed unfit in Section 2.1, so we clarified: our control vector is x_0 the system initial state, and not x as we wrote in the first version of our paper. All occurrences of x have been replaced accordingly in Section 2.1, so that the use of $\mathcal{G}$ is now fit.

7) P5,l23: Precise the page/section number in Ambrosio et al. (2008).

It is more easily accessible in Benamou and Brenier, 2000, so we actually changed the reference.

8) p 10, l14-18: Remind the equation number associated with the cost function and gradient. L^2 cost function is related with Eq.(2), Wasserstein cost function with Eq.(14), and the iteration steps are deduced from Eq.(18). **Ok**

9) P9,l17:write the push-forward for a given $x \in \Omega$ as $\rho_1[T(x)]|det\nabla T_x| = \rho_0(x)$.

This remark has been removed, following Referee's 2 comment. See Section 3.3 where the Monge-Ampere terminology of OT (with a transport map T) has been removed to only deal with the Benamou and Brenier formulation (with v).

10) P10, l19: "α^n is chosen as optimal": explain how it is computed, and provide an appropriate reference.

We specified that α^n is found using a line search algorithm. It is therefore not strictly optimal but approximately optimal.