

Interactive comment on “Optimal Transport for Variational Data Assimilation” by Nelson Feyeux et al.

Anonymous Referee #2

Received and published: 19 September 2017

This is a very interesting paper, which introduces the use of optimal transport, and its metrics, in the world of data assimilation. The layout of the paper is very clear and appealing. For these reasons I strongly recommend the publication of this paper. However, the manuscript could benefit from the following remarks and suggestions:

- The English could and should be significantly improved.
- Clarifications are needed now and then, especially to help the reader who has little acquaintance with optimal transport.
- There is a bit of a contradiction in the willingness to introduce, or not, the transference plan view on optimal transport. This should be clarified.

C1

- The authors could get rid of the mathematical remark style. In my opinion, it is not suited for NPG and is detrimental to the clarity of text. For the present manuscript, all the remarks can easily be naturally embedded in the text.
- For the sake of clarity, you should precisely define y_i^{obs} and x_0^b as mathematical objects.
- A recurrent question in data assimilation, which I believe many of your readers will have is: is there any probabilistic interpretation of the cost function defined with the Wasserstein distance? This is worth discussing it briefly.
- The paper would benefit from a more detailed discussion of the experiments, possibly another one with noise in the observation. There is room for it.

Specific remarks, in connection, or not, to the previous remarks are:

1. Title: why the capital letters in the title?
2. p.1, l.5: "With appropriate choices...": of what? Unclear.
3. p.1, l.6: "Optimal-transport-based optimization..." → "Optimal transport-based optimization..."
4. p.1, l.6-7: "...to preserve the geometrical properties of the estimated initial condition.": this statement is too mysterious for an abstract. You should be more explicit.
5. p.1, l.13: "to use so-called data assimilation methods" → "to use the so-called data assimilation methods"
6. p.1, l.14: It is uncanny that the third author is reluctant to cite her own brand new book on data assimilation.

C2

7. p.1, l.14-15: "They aim at finding either the initial/boundary conditions or some parameters of a numerical model.": not only! They can be used for parameter estimation, reanalysis, etc.
8. p.1, l.18: "comparison between the observations and their model counterparts.": a mathematical expression called the innovation in data assimilation.
9. p.1, l.19: "unperfect" → "imperfect"
10. p.1, l.22: "More recently an hybrid of both approaches..." → "More recently hybrids of both approaches..."
11. p.2, l.1-2: "model counterparts. A Tikhonov regularization is also added and so the distance between the control vector and a background state carrying the a priori information is added in the cost function.": needs to be rephrased. It could be instead: "A Tikhonov regularization term is also added to the cost function as a distance between the control vector and a background state carrying the a priori information."
12. p.2, l.4: "aims to reach" → "aims at reaching"
13. p.2, l.4: "are smallest as possible." → "are as small as possible."
14. p.2, l.13: "...the desired localization." → "...the desired location."
15. p.2, l.17: "...has been founded by Monge..." → "...has been pioneered by Monge..."
16. p.2, l.19: I would remove "quickly".
17. p.3, l.1: "from pure mathematical analysis" → "from pure mathematical analysis on Riemannian spaces"

C3

18. p.3, l.8: "...Wasserstein distance is to compare..." → "...Wasserstein distance to compare..."
19. p.3, l.9: "data assimilation Actual use of optimal transport" → "data assimilation. Actual use of optimal transport". Better, you could start a new paragraph with "Actual use..."
20. p.3, l.15: "This particularly subtle mathematical consideration is indeed crucial for the algorithm..." → "This particularly subtle mathematical considerations are indeed crucial for the algorithm..."
21. p.3, l.18: "...methods but it largely exceeds..." → "...methods, which largely exceeds..."
22. p.3, l.21: "required for the sequel" → "required in the following"
23. p.3, l.23-24: "Section 4 numerical illustrations are presented, choices for the gradients and the optimization methods are compared.": could be improved. Please rephrase.
24. p.3, l.24: "...and solutions proposed." → "...and solutions will be proposed."; the ellipsis could be avoided here.
25. p.3, l.26: "The section..." → "This section..."
26. p.3, l.27: "...materials...": principles?, facts?, properties?
27. p.3, l.28: "...production." → "...contribution."
28. p.4: You could mention that the Euclidean distances are local metrics, as opposed to the Wasserstein distance.
29. p.4, l.6: "...term x_b which contains..." → "...term x_b , which contains..."

C4

30. p.4, l.6: "The actual cost function then writes..." → "The actual cost function then reads..."
31. p.4, l.15: "[0, 1]" → "the interval [0, 1]" since the notation is not really universal.
32. p.5, section 2.2.2: explain that the time t is fictitious, or you will puzzle many readers.
33. p.5, l.9-11: Actually, I don't believe this is a necessary condition. There could non-zero fluxes of probability with a global balanced budget; see for instance Farchi et al. (2016).
34. p.5, l.18: Use `\citep` for the citation to Benamou and Brenier (2000).
35. p.5, l.21: "A remarkable point..." → "A remarkable property..."
36. p.5, l. 22: Use `\citep` for the citation to Ambrosio et al. (2008).
37. p.6, l.4-5: "...like the primal-dual Papadakis et al. (2014) or the semi-discrete Mérigot (2011).": I would be thrilled in meeting the primal dual Papadakis or discussing with the semi-discrete Mérigot... Please rephrase.
38. p.6, l.7: "... the scalar product choice conditions the gradient value." → "... the scalar product choice is used to define the gradient value."
39. p.6, l.11: "...shall formally be defined by..." → "...is formally defined by..."
40. p.6, l.11: "(cf. Otto (2001))": use `\citep[] [] {}`.
41. p.6, Eq.(9): you probably should mention the set to which the Kantorovitch potential belongs.
42. p.6, l.15: This is not a proper sentence; you could merge it with the previous one.

C5

43. p.6, l.23: "First we will consider..." → "First, we will consider..."
44. p.6, l.24: "Second we will investigate..." → "Second, we will investigate..."
45. p.6, l.24: "...we will investigate the role of the scalar product choice as well as the gradient descent method..." → "...we will discuss the choice of the scalar product as well as the choice of the gradient descent method..."
46. p.7, l.5-6: another example, more accessible to the NPG readership, is the distance built in Farchi et al. (2016).
47. p.7, l.11: "...belonging respectively to $\mathcal{P}(\Omega)$ and $\mathcal{P}(\Omega_0)$." → "...belonging to $\mathcal{P}(\Omega)$ and $\mathcal{P}(\Omega_0)$, respectively."
48. p.7, l.19: The scalar product is not unique (and as a consequence the gradient), but there is a natural one induced by the norm used in the cost function (here Wasserstein's). This could be mentioned, as the statement could be slightly puzzling for the reader.
49. p.7, l.26: It is not clear at this stage why you would use the \mathcal{L}^2 inner product.
50. p.8, l.24-25: Two "thus" in a row.
51. p.9, l.19: "...we will use after." → "...we will use in the following."
52. p.9, l.14-24: You mentioned p.5, l.18-19 that the definition of optimal transport based on transference map is out of scope; and I am fine with it. I even think it was a clever choice. But, here, you finally use it and that seems important. This is quite frustrating for the reader, especially those who have little knowledge on optimal transport.
53. p.10, l.5: ", but results are still satisfactory.": Please remove the statement. It does not make sense to give the conclusion beforehand.

C6

54. p.10, l.8-13: Why not consider, in addition, a case with observation noise; you perturb the Gaussian parameters of the observation, which would be similar to some bias in satellite observation.
55. p.10, l.19: "is chosen a optimal": vague, please be more specific.
56. p.11, Eq.(26): I would explicitly write the wind field in the equation even if it is uniformly equal to 1.
57. p.12, l.4: "...gaussians..." → "...Gaussians..."
58. p.12, l.1: "The analyses $\rho_0^{a,W,2}$ and $\rho_0^{a,W,\#}$ are different even if they arise from the same cost function \mathcal{J}_W , which highlights the need for a well-suited scalar-product.": that is one of the most interesting point of the experiment, but your comment is too short. You must elaborate. One would expect the numerical solutions to be the same, right? unless there is a convergence issue, which much be analysed and discussed and would fit nicely with what was laid in section 3.2.
59. p.12, l.18: "Shallow-Water" → "shallow-water"
60. p.12, l.26: "Thanks to the wisdom gained..." → "Thanks to the experience gained...": My wisdom told me that norm-induced scalar product was the best one from the very beginning.
61. p.13, Figure.4: please plot the observations, like you did for the first experiment.
62. p.13, l.9: "...badly..." → "...poorly..."
63. p.12, l.12: Please avoid inverting subject and verb as this is much less frequent in English than in French.

C7

References

- Ambrosio, L., Gigli, N., Savaré, G., 2008. Gradient flows: in metric spaces and in the space of probability measures. Springer Science & Business Media.
- Benamou, J.D., Brenier, Y., 2000. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik* 84, 375–393.
- Farchi, A., Bocquet, M., Roustan, Y., Mathieu, A., Quérel, A., 2016. Using the Wasserstein distance to compare fields of pollutants: application to the radionuclide atmospheric dispersion of the Fukushima-Daiichi accident. *Tellus B* 68, 31682. doi:10.3402/tellusb.v68.31682.

Interactive comment on Nonlin. Processes Geophys. Discuss., <https://doi.org/10.5194/npg-2017-46>, 2017.

C8