

# Review – Convergence Rates of Gaussian ODE Filters

by Hans Kersting, T.J. Sullivan and Philipp Hennig

The authors present convergence results for a numerical method introduced in [M. Schober et al., *Probabilistic ODE solvers with Runge-Kutta means*, 2014] and further developed in [H. Kersting and P. Hennig, *Active uncertainty calibration in Bayesian ODE solvers*, 2016]. The main idea underlying this numerical scheme is to exploit techniques inherited from Kalman filters to establish a Gaussian measure over the solution of initial value problems and its derivatives. Given a Gaussian process prior on the solution of the initial value problem, the output is a Gaussian posterior obtained by repeated evaluation of the right hand side (or, in the phrasing of the manuscript, the data) and by the application of Bayes' rule. This method guarantees fast computations, fully deterministic, to output a measure over numerical solutions and thus fits perfectly in the emerging framework of probabilistic numerical methods.

The main contribution of this work is proving a convergence result of the filtering approach, i.e., that the mean of the Gaussian measure tends to the true solution for the time step  $h \rightarrow 0$  while the variance goes to zero in the same limit. Convergence is analysed in local and global terms. However, while in the local case a general framework is considered, where  $q$  derivatives of the numerical solution are modelled and convergence is of order  $q$ , the global analysis is limited to the case  $q = 1$ , i.e., only the solution itself and its first derivative are taken into account. Nonetheless, already proving a first order result is technical and manifestly required the authors a big effort and lengthy computations.

My main remarks for the authors are the following.

- (i) The authors should stress more the advantages of the filtering approach vs. the probabilistic methods presented in Abdulle and Garegnani (2018), Conrad et al. (2017), Lie et al. (2017) (see references of the reviewed paper). In particular, the theoretical effort which seems to be necessary to prove a global convergence of order 1 makes it dubious that a coherent uncertainty quantification of methods of higher order could be given in the filtering framework. Hence, the competitiveness of the filtering approach in the family of probabilistic numerical methods is unclear, at least from a theoretical point of view. Moreover, for nonlinear ordinary differential equations, where the solution could exhibit a chaotic behaviour, a Gaussian output seem to be inappropriate and overly restrictive.
- (ii) Given the novelty of the convergence properties which are presented in this work, I believe that the authors should show numerical experiments for an empirical verification of these results. In particular, since no global convergence of order  $q$  is proved theoretically, an experiment showing that this global order is attained in practice should be provided, at least to support the discussion of Section 8.
- (iii) In Section 8, the filtering method is said to be employable in the frame of inverse problems to enhance their solution qualitatively. The methods proposed in Abdulle and Garegnani, Conrad et al., Lie et al. seem to fit perfectly in pseudo-marginal MCMC methods and there is empirical evidence about their usefulness in this framework. Hence, the authors should clarify how it would be possible to integrate the filter into an inverse problem solution.
- (iv) I believe that the results of local convergence would highly benefit in terms of clarity if the two cases IBM and IOUP would be separated, as the proof for the case IBM is direct when the (more complicated) results for IOUP are given and could therefore be skipped. In this way, the heavy notation

$$\int \cdot \bigg\}$$

could be abandoned.

- (v) The choice of discussing global convergence only for IBM seems too restrictive and makes the usefulness of a result of local convergence for IOUP dubious, even in light of the discussion in Section 8. Is it possible to show this at least empirically by means of numerical experiments?
- (vi) Section 6.2-3 lack any presentation/motivation/explanation of the presented theoretical results, as well as their late usefulness in the proof of Theorem 6.7.
- (vii) In Section 8, the authors state that “the rates of  $h^{q+1}$  (local) and  $h^q$  (global) are optimal”. Leaving out the fact that the global rate of  $h^q$  has been proved only in the case of  $q = 1$ , the authors should clarify in which sense is optimality thought of. For example, Runge-Kutta methods on  $s$  stages can provide an (optimal) order of convergence of  $2s$  in case the quadrature points are well chosen (Gauss sequences), going well beyond the equality  $s$  (stages)  $= s$  (order).
- (viii) In general, I believe that the authors should invest some energy into a thorough revision in terms of text quality and overall readability of their manuscript.

Moreover, a list (most likely incomplete) of necessary modifications I found is the following

- Lines 21–22, parentheses do not look necessary,
- Line 113, consistency between *Section* and *section*,
- Subsection 1.4 (typo at line 131),
- The notation  $d\mathbf{X}_{j,t}$  at line 150 is not introduced in the text and seems inconsistent with the following notation  $m_j(t)$  and  $P_j(t)$ ,
- Lines 170 and 173, consistency between  $\mathbb{1}_{i \leq j}$  and  $\mathbb{1}_{j \leq i}$ ,
- Lines 255–261, introduction of bold symbols,
- Lines 300–301, the symbol  $\Delta^{(i)}$  does not seem to belong to this sentence,
- Lines 306–308, a clear and rigorous definition of state misalignment should be provided,
- Theorem 5.2, the notation  $\varepsilon$  should be recalled (only introduced in the text at the beginning of Section 4),
- Proposition 6.2, the notation  $\wedge = \min$  and  $\vee = \max$  should be specified in the text,
- Line 461, I believe it should be
 
$$P_{11}^\infty = \lim_{n \rightarrow \infty} P_{11}(nh),$$
- Line 491, misalignment of equation and text,
- Line 586–587, phrasing,
- Proof of Lemma 6.4, lack of notation consistency between  $m^{-\cdot(0)}$  of e.g. (2.10) and  $m^{-(0)}$ ,
- Equation (6.33),  $m^{(-1)}$  should be  $m^{-(1)}$  or  $m^{-\cdot(1)}$  (see comment above),
- Equation (6.33) and (6.34),  $r(nh)$  should exit the norm in absolute value (does not spoil the overall correctness of the proof),
- Line 625, reference to Proof 14 (proofs are unnumbered),
- Line 698–699, phrasing and why half the rate (and not the same rate),
- Line 704, misalignment of equation and text,
- Line 728, phrasing.