

DA/Bayes 推定の理論的 review(メモ)

竹田航太

2023 年 3 月 14 日

目次

1	数学的定式化	1
1.1	設定	2
1.2	基本的な目標	3
1.3	示すべきこと	3
2	結果	4
2.1	Bayes	4
2.2	(MLE Consistency)	5
2.3	DA	5
2.4	Model	8
3	解析手法	8
4	研究者	8
4.1	David Kelly	8
4.2	Xin T. Tong	8
4.3	Jana de Wiljes	9

1 数学的定式化

データ同化やその元となる Bayes 推定の数学的定式化と示すべき事項について整理する。ベイズの定理に従って観測値から逆推定を行う。観測作用素とノイズの仮定から尤度を定め、事前分布を修正し事後分布が定義される。このような事後分布がどれくらい真の状態を推定できているかを評価する必要がある。以上までは理想的な分布の話であり、実際のデータ同化アルゴリズムは理論的な事後分布を何らかの方法で近似し、計算機上で実現する。この近似に対する評価も

同時に必要となる．空間・時間離散化によるものや分布の近似からくる誤差が考えられる．また，このような分布に関する評価を用いて，平均値などによる点推定の誤差評価も求められる．

1.1 設定

状態空間として，Hilbert 空間 \mathcal{V} をとる．モデルとして力学系を考える^{*1}．

$$\frac{du}{dt} = f(u). \quad (1.1)$$

真の状態 $(u^\dagger(t))_{t \geq 0}$ は初期値 $u_0^\dagger \in \mathcal{V}$ から (1.1) による時間発展で得られるとする．モデル (1.1) に従う時間発展が 1-パラメータ半群 $\Psi_t : \mathcal{V} \rightarrow \mathcal{V}$ で書けると仮定し，時間 $h > 0$ の時間発展を $\Psi = \Psi_h$ とおくと，以下の離散力学系を考えることもできる．

$$u_j = \Psi(u_{j-1}) \quad (1.2)$$

Ψ_h が明示的に書けるとは限らないので，実用上は時間方向に離散化したものを Ψ_h とすることが多い．また，ノイズを含む力学系を考える場合もある．

$$u_j = \Psi(u_{j-1}) + \xi_j, \quad \xi_j \sim N(0, \Sigma) \quad (1.3)$$

ただし，ノイズ $\xi_n \in \mathcal{V}$ は i.i.d. な Hilbert 空間上の Gauss ノイズであり，Covariance 作用素 $\Sigma : \mathcal{V} \rightarrow \mathcal{V}$ は対称・半正定値とする．

観測時間間隔 $h > 0$ と $d \in \mathbb{N}$ に対して， \mathbb{R}^d 上の観測値が観測作用素 $\mathcal{H} : \mathcal{V} \rightarrow \mathbb{R}^d$ を通して得られる．

$$y_j = \mathcal{H}(u^\dagger(hj)) + \eta_j, \quad \eta_j \sim N(0, \Gamma)$$

ここで， $\eta_j \in \mathbb{R}^d$ は観測ノイズである．ここでは，Covariance $\Gamma \in \mathbb{R}^{d \times d}$ を持つ Gauss ノイズとしたが一般にはそのような分布に従わない．

表 1: モデルの設定

時間離散		時間連続	空間離散		空間連続
ノイズなし	離散力学系	連続力学系	ノイズなし	半離散力学系	無限次元力学系
ノイズあり	Markov 連鎖	SDE	ノイズあり	SDE	SPDE

(a) ODE.

(b) PDE. 時間は連続とする.

^{*1} PDE は無限次元の力学系として考えるか，空間離散化をして有限次元に帰着する．

1.2 基本的な目標

データ同化の目的は $J \in \mathbb{N}$ 個の観測値 $(y_j)_{j=1}^J$ から、初期値に対する推定を行うことである。もしくは、観測値が得られている最終時刻より先の時刻での状態を推定することである。ノイズのないモデルの場合は初期値が分かれば後の時刻の全ての状態が決まる。気象予報などでは Ψ_h の計算に時間がかかり、時間と共に観測値が増えていくため、各観測時刻で状態に対する推定値 \hat{u}_j が得られる方が望ましい。また、推定値に加えて推定の不確実性が分布として得られるとより良い。このような目的を達成するため、初期値に関する推定分布 $\mu_0(du_0)$ を適当に与え、モデル Ψ^{*2} や観測に仮定を置いて推定分布の列 $(\mu_j(du_j))_{j=1,2,\dots}$ を計算することがデータ同化の基本的な問題設定となる。

Remark 1.1. 最初に与えるもの

- 推定初期値 $\mu_0(du_0)$.
- モデル Ψ . ノイズの情報もあれば
- 観測作用素 \mathcal{H} .
- 観測分布. 観測値から推定できているとする.
- 観測値 $(y_j)_{j=1}^J$. J がどんどん大きくなっていく場合もある.

1.3 示すべきこと

1.3.1 Wellposedness

事後分布推定問題における解の存在と一意性とパラメータ（観測データ）に対する連続性.

1.3.2 Posterior consistency

十分なデータが得られれば、真の状態 v^\dagger を recover できること。データ数が ∞ , 観測ノイズが 0 の極限において事後分布が真の状態にサポートを持つ Dirac 測度に収束することが保証されるべき。 $Y_J = (y_n)_{n=1}^J$,

$$d(\mu^{(Y_J, \Gamma)}, \delta_{u^\dagger}) \rightarrow 0 ? \quad (J \rightarrow \infty, tr(\Gamma) \rightarrow 0).$$

1.3.3 Bayesian Quality Assessment

理論的な事後分布 μ の個別のアルゴリズムによる近似 μ_{approx} がどれくらい近いかを評価すること。(離散化パラメータに対する連続性とも言える?)

^{*2} 真の時間発展 Ψ は一般には未知なので代わりに推定及び近似した $\hat{\Psi}$ とおき、誤差について議論すべき場合もある。

$$d(\mu_{approx}, \mu) \leq ?.$$

1.3.4 Signal Quality Assessment

信号推定（点推定）問題において推定値 \hat{u} が真の状態 u^\dagger をどれくらい近いかを評価すること。まず、データが真の状態を recover するのに十分な情報を持っているか (MLE consistency とも言える?) を評価し、また、個別のアルゴリズムがどの程度情報を取り出せるかを評価する必要がある。

$$\limsup_{j \rightarrow \infty} \mathbb{E}[\|\hat{u}_j - u^\dagger(jh)\|_V] \leq ?.$$

1.3.5 Ergodicity

PF 型の推定分布が一意的な invariant measure に収束するか、それがどれくらい速いかを評価する。

1.3.6 (model の Wellposedness 等)

モデルの誤差発達の評価がデータ同化の解析に必要である。そのために、モデル方程式の解や global attractor の存在などについても調べる必要がある。

2 結果

2.1 Bayes

[1] 定式化の際には、尤度やポテンシャル (= $-\log$ 尤度) に対して条件が課せられる。これは観測関数とノイズの設計に対する条件とも言える。

- (1) Standard assumption[1]: ポテンシャルに関する有界性や Lipschitz 性。Wellposedness に必要。
- (2) Identifiable[1]: 推定パラメータに対する尤度の”単射”性。
- (3) Regularity assumption[1]: 尤度に対する可微分性、可積分性の条件。最尤推定と比較する文脈 (Consistency) において用いられる。

2.1.1 Wellposedness

無限次元 (1)generalized Bayes(H: Banach to Euclid): これで事後分布の形が決まる。(2)well-defined posterior: 事後分布が well-defined であることが示される。(3)Locally Lipschitz: 観測

データに対する Hellinger 距離での Lipschitz 性.

2.1.2 Posterior Consistency

Bernstein–von Mises(Bayesian)

2.2 (MLE Consistency)

[1] (1)Consistency of the MLE, (2)Local asymptotic normality of the MLE

2.3 DA

2.3.1 真の Filtering 分布 (時間離散)

Theorem 2.1 (Locally Lipschitz w.r.t. obs, Theorem 2.15 in [2]). $\mathcal{V} = R^l$ とする. (1.3) に対する *smoothing problem* を考える. *path* 空間 \mathcal{V}^J 上の事前分布 μ_0 に対し $\mathbb{E}^{\mu_0}[\sum_{n=1}^J(1 + |\mathcal{H}(u_n)|^2)] < \infty$ とする. このとき, 任意の $r > 0$ に対して, ある $c = c(r)$ が存在して

$$d_{Hel}(\mu^y, \mu^{y'}) \leq c|y - y'|, \quad \forall |y|, |y'| \leq r.$$

ただし, μ^y と $\mu^{y'}$ は \mathcal{V}^J 上の事後分布. ([1] の *Locally Lip* と比較すべき)

$$J(u) = \frac{1}{2}|u_0 - m_0|_{C_0^{-1}}^2 + \frac{1}{2} \sum_{j=1}^J |u_j - \Psi(u_{j-1})|_{\Sigma^{-1}}^2,$$

$$\Phi(u; y) = \frac{1}{2} \sum_{j=1}^J |y_j - \Psi(u_j)|_{\Gamma^{-1}}^2,$$

$$I(u; y) = J(u) + \Phi(u; y).$$

MAP の wellposedness

Theorem 2.2 (MAP/4DVar と事後分布の関係 (Theorem 3.10, 12)). $\mathcal{V} = R^l$ とする. *stochastic model*(resp. *deterministic model*) に対して, $\Psi \in C^1$, $\mathcal{H} \in C^1$ のとき, $I(\cdot; y)$ (resp. $I_{det}(\cdot; y)$) は少なくとも一つの最小化点 $u^* \in \mathcal{V}^J$ (resp. $u_0^* \in \mathcal{V}$) を持ち, 事後分布は u^* で最大となる.

2.3.2 KF

1 次元 KF を以下の model と観測について考える.

$$\Psi(u) = \lambda u, \mathcal{H}(u) = u, \quad \Sigma = \sigma^2, \Gamma = \gamma^2.$$

初期推定値 $m_0 \in \mathbb{R}$, $c_0 > 0$ で表す. Covariance c_j は表??の挙動を示す.

表 2: 1 次元 KF の Cov の長時間挙動, Table 4.1 in [2]

	Limiting covariance for $\sigma^2 = 0$	Limiting covariance for $\sigma^2 > 0$
$ \lambda < 1$	$c_j \rightarrow 0$ (exponentially)	$c_j \rightarrow c_+^* = \mathcal{O}(\gamma^2)$ (exponentially)
$ \lambda = 1$	$c_j \rightarrow 0$ (algebraically)	$c_j \rightarrow c_+^* = \mathcal{O}(\gamma^2)$ (exponentially)
$ \lambda > 1$	$c_j \rightarrow c_-^* = \mathcal{O}(\gamma^2)$ (exponentially)	$c_j \rightarrow c_+^* = \mathcal{O}(\gamma^2)$ (exponentially)

2.3.3 Particle Filter

真の filtering 分布を実際のアルゴリズムがどの程度近似できているかを評価する。Filtering step, Forecast step と sampling step をそれぞれ以下のように表す。^{*3}

$$\begin{aligned}
 (L_j \mu) du &= \frac{g_j(u) \mu(du)}{\int g_j(u) \mu(du)}, \\
 (P \mu)(du) &= \int p(v, du) \mu(dv), \\
 (S^N \mu)(du) &= \frac{1}{N} \sum_{n=1}^N \delta_{u^{(n)}}(du), \quad u^{(n)} \sim \mu \text{ i.i.d..}
 \end{aligned}$$

ただし, $g_j(u_{j+1}) \propto \mathbb{P}(y_{j+1}|u_{j+1})$, $p(v, du)$ は (1.3) に対応する Markov kernel, $N \in \mathbb{N}$ はサンプルサイズ. PF で得られる filtering 分布 μ_j^N は以下を満たす.

$$\mu_{j+1}^N = L_j S^N P \mu_j^N, \quad \mu_0^N = N(m_0, C_0).$$

真の filtering 分布 μ_j は以下を満たす.

$$\mu_{j+1} = L_j P \mu_j, \quad \mu_0 = N(m_0, C_0).$$

Bayesian Quality Assessment

Theorem 2.3 (PF の離散化誤差, Theorem 4.5 in [2]). ある $\kappa \in (0, 1]$ が存在して任意の $u \in \mathcal{V}, j \in \mathbb{N}$ で $\kappa \leq g_j(u) \leq \kappa^{-1}$ が成り立つとする. このとき, 以下が成り立つ.

$$d_{TV}(\mu_J^N, \mu_J) \leq \sum_{j=1}^J (2\kappa^{-2})^j \frac{1}{\sqrt{N}}.$$

OPF の誤差解析と Ergodicity[3, Kelly-Stuart-Majda 2019].

^{*3} この表記は一般に使える.

2.3.4 3DVar

Theorem 2.4 ((部分観測)3DVar の stability, Theorem 4.10 in [2]). $\mathcal{V} = R^l$ とする. 観測が線形 $H : \mathcal{V} \rightarrow \mathbb{R}^d$ とする. 観測誤差が有界とする (*i.e.*, $\sup_{j \in \mathbb{N}} |\epsilon_j| = \epsilon$). *deterministic* な *model* を考え, $a < 1$ に対して, $(I - KH)\Psi : \mathcal{V} \rightarrow \mathcal{V}$ が a -global Lipshitz となるように *Background Covariance* B をとるとする. このとき, ある $c > 0$ が存在して以下が成り立つ.

$$\limsup_{j \rightarrow \infty} \|m_j - u_j^\dagger\| \leq \frac{c}{1-a} \epsilon.$$

$\frac{c}{1-a} < 1$ と取れるとき, 3DVar で観測より良い精度で推定できていることになる. 変数の依存関係は $K = K(B, \Gamma), B = B(\sigma, \Psi), a = a(K, H, \Psi), c = c(K)$.

Theorem 2.5. [4, Law-Shukla-Stuart 2014] Lorenz63, $H = (1, 0, 0)$, 離散・連続 3DVar の誤差解析.

Theorem 2.6. [5, Law-Alonso-Shukla-Stuart 2016] Lorenz96, 部分線形観測, 3DVar の誤差解析

2.3.5 EnKF

Theorem 2.7 (PO(連続・離散) の誤差解析, Theorem 4.2 in [6]). 誤差評価がわかっている *deterministic model* とノイズ $N(0, \gamma^2 I)$ を持つ完全観測 $\mathcal{H} = I$ を仮定する. PO 法の *ensemble* 数を $N \in \mathbb{N}$ とする. $e_j^{(n)} = u_j^\dagger - u_j$ と書く. このとき, 各 *particle* の *SE* の期待値は

$$\mathbb{E}[|e_j^{(n)}|^2] \leq e^{2\beta h j} \mathbb{E}[|e_0^{(n)}|^2] + 2N\gamma^2 \left(\frac{e^{2\beta h j} - 1}{e^{2\beta h} - 1} \right), \quad n = 1, \dots, N, j \in \mathbb{N}.$$

さらに, *addtive inflation* $C \rightarrow C + \alpha^2 I$ ($\alpha > 0$) を考えると, $\theta = \frac{\gamma^2}{\gamma^2 + \alpha^2} e^{2\beta h}$ に対して,

$$\mathbb{E}[|e_j^{(n)}|^2] \leq \theta^j \mathbb{E}[|e_0^{(n)}|^2] + 2N\gamma^2 \left(\frac{1 - \theta^j}{1 - \theta} \right), \quad n = 1, \dots, N, j \in \mathbb{N}.$$

$\theta < 1$ のとき, 右辺は $j \rightarrow \infty$ で $\frac{2N\gamma^2}{1-\theta}$ に収束する.

[7, Tong-Majda-Kelly 2016] は EnKF(PO) と ESKF を扱い, 線形観測を考えた. エネルギー不等式と Lyapunov 関数を model や filtering の特徴づけに用いて, filtering 後のエネルギーの有界性を示し, filtering の安定性とした. (EnKF に付随する Markov chain の Lyapunov 関数を構成するために model のエネルギー不等式を利用した.) EnKF では真の filtering 分布への収束は言えないが, 縮小的であり一意な "invariant measure" に収束することを保証した.

2.3.6 EnKBF

EnKBF(時間連続) の誤差解析 [8, Wiljes-Reich-Stannat 2018] は完全観測, small noise などの条件で解の存在や誤差解析をしている. サンプルサイズ無限大での Covariance の固有値の bound も与えている. Localized EnKBF(時間連続 LEFKT) の誤差解析 [9, Wiljes-Tong 2020].

2.3.7 Filter divergence

[10, Kelly-Majda-Tong 2015]

2.4 Model

モデルの解析について [11, 5, 6] が Lorenz63, 96 や 2 次元周期境界 Navier-Stokes 方程式の解の存在や誤差発達について議論した. [12] では正則化 3 次元 Navier-Stokes 方程式に対して, 同様に線型項や非線型項の評価が得られている.

3 解析手法

Gronwall の不等式, エネルギー不等式, Markov process/SDE の解析 (Coupling, Lyapunov 関数, Minorisation), 行列の固有値/スペクトル解析, Random Matrix,

4 研究者

4.1 David Kelly

google scholar: https://scholar.google.com/citations?hl=ja&user=c-dF0dIAAAAJ&view_op=list_works&sortby=pubdate

4.2 Xin T. Tong

National University of Singapore <https://sites.google.com/view/xintongthomson/home>

キーワード: DA, Sampling, Optimization, Nonlinear SDE

研究紹介スライドがいろいろある. <https://sites.google.com/view/xintongthomson/slides>

4.3 Jana de Wiljes

Potsdam 大学. Tong や Reich と共著がある. <https://www.math.uni-potsdam.de/~jdewiljes/>

参考文献

- [1] Timothy John Sullivan. *Introduction to uncertainty quantification*, volume 63 of *Texts in Applied Mathematics*. Springer, 2015.
- [2] K. J. H. Law, A. M. Stuart, and K. C. Zygalakis. *Data Assimilation: A Mathematical Introduction*. Springer, 2015.
- [3] David Kelly and Andrew M. Stuart. Ergodicity and accuracy of optimal particle filters for bayesian data assimilation. *CHINESE ANNALS OF MATHEMATICS SERIES B*, 40(5):811–842, SEP 2019.
- [4] Kody Law, Abhishek Shukla, and Andrew Stuart. Analysis of the 3dvar filter for the partially observed lorenz’63 model. *DISCRETE AND CONTINUOUS DYNAMICAL SYSTEMS*, 34(3, SI):1061–1078, MAR 2014.
- [5] K. J. H. Law, D. Sanz-Alonso, A. Shukla, and A. M. Stuart. Filter accuracy for the lorenz 96 model: Fixed versus adaptive observation operators. *PHYSICA D-NONLINEAR PHENOMENA*, 325:1–13, JUN 15 2016.
- [6] D. T. B. Kelly, K. J. H. Law, and A. M. Stuart. Well-posedness and accuracy of the ensemble kalman filter in discrete and continuous time. *NONLINEARITY*, 27(10):2579–2603, OCT 2014.
- [7] Xin T Tong, Andrew J Majda, and David Kelly. Nonlinear stability and ergodicity of ensemble based kalman filters. *Nonlinearity*, 29(2):657, jan 2016.
- [8] Jana de Wiljes, Sebastian Reich, and Wilhelm Stannat. Long-time stability and accuracy of the ensemble kalman–bucy filter for fully observed processes and small measurement noise. *SIAM Journal on Applied Dynamical Systems*, 17(2):1152–1181, 2018.
- [9] Jana de Wiljes and Xin T Tong. Analysis of a localised nonlinear ensemble kalman bucy filter with complete and accurate observations. *Nonlinearity*, 33(9):4752–4782, jul 2020.
- [10] David Kelly, Andrew J. Majda, and Xin T. Tong. Concrete ensemble kalman filters with rigorous catastrophic filter divergence. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 112(34):10589–10594, AUG 25 2015.

- [11] Kevin Hayden, Eric Olson, and Edriss S. Titi. Discrete data assimilation in the lorenz and 2d navier-stokes equations. *PHYSICA D-NONLINEAR PHENOMENA*, 240(18):1416–1425, SEP 1 2011.
- [12] C. Foias, D. D. Holm, and E. S. Titi. The three dimensional viscous camassa-holm equations, and their relation to the navier-stokes equations and turbulence theory. *Journal of Dynamics and Differential Equations*, 14(1), 2001.
- [13] Richard Nickl. Statistical theory. *Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge*, 2013.
- [14] 知之 樋口, 玄太 上野, 慎也 中野, 和幸 中村, and 亮 吉田. データ同化入門 : 次世代のシミュレーション技術. Number 6 in シリーズ予測と発見の科学. 朝倉書店, 2011.
- [15] Marco A Iglesias, Kody J H Law, and Andrew M Stuart. Ensemble kalman methods for inverse problems. *Inverse Problems*, 29(4):045001, mar 2013.
- [16] Andrew J. Majda and Xin T. Tong. Rigorous accuracy and robustness analysis for two-scale reduced random kalman filters in high dimensions, 2016.
- [17] Sean P. Meyn and Richard L. Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2005.