数値解析と確率・統計による不確実性定量化

東京大学大学院 情報理工学系研究科・理化学研究所 脳神経科学研究センター 松田 孟留 *
Takeru Matsuda
Graduate School of Information Science and Technology, The University of Tokyo
RIKEN Center for Brain Science
大阪大学サイバーメディアセンター 宮武 勇登 †
Yuto Miyatake
Cybermedia Center, Osaka University

近年，さまざまな分野において膨大なデータを得ることが可能となってきており，また計算機の高性能化に伴って数値シミュレーションの大規模化も著しい．こうした動向を反映して，数値解析と確率・統計を融合した不確実性定量化 (uncertainty quantification) に関する研究の重要性が高まっている．本稿の前半では，不確実性定量化のためにベイズ統計の観点から数値計算を捉え直す Probabilistic Numerics という分野について紹介し，この分野の今後の展望を述べる．本稿の後半では，筆者らの最近の研究のうち，Probabilistic Numerics の着想に基づいた常微分方程式のパラメータ推定手法 [25] について簡単に紹介する．

1　数値解析と確率・統計のつながり

本研究集会の目的として「学問全体が急速に第四の科学，すなわちデータ駆動型科学へと転回し新しいトピックが湧き起こるいま，計算を司る数値解析学が果たせる新しい可能性について探究」,「さらにそれらを通じて，第四の科学時代に諸科学分野を結ぶ基礎学問としての数値解析学のあり方を改めて問い直すこと」が挙げられていた．本章では，数値解析と確率・統計のつながりという観点からこれらの課題について考察してみたい．

1.1　不確実性定量化 (uncertainty quantification)

近年，さまざまな分野において膨大なデータを得ることが可能となり，データをもとに現象の理解・予測・制御を行うデータ駆動型科学の重要性が高まっている．データ解析の方法論は統計学の分野で古くから研究されており，今後も統計学的手法はデータ駆動型科学における基本的な道具として用いられると考えられる．統計学ではデータが確率的に生成されるという仮定をおき，データが従う確率分布を「統計モデル」の形に表現する．これら統計モデルをデータに当てはめることで，背後にある現象を「データに語らせる」ことが大きな目標である．一方，計算機の高性能化に伴って，現象を記述する物理モデルを精密に解く数値シミュレーションの大規模化も著しい．大規模数値シミュレーションと大規模観測データを融合するデータ同化の手法は，地球科学をはじめとする多くの分野で活用されている．こうした動向を反映して，数値解析と確率・統計を融合した不確実性定量化 (uncertainty quantification) に関する研究に注目が集まっている．以下は，2013 年に SIAM が ASA(アメリカ統計学会) と共同で創刊した SIAM/ASA Journal on Uncertainty Quantification の趣旨に関する文章である (ウェブサイトから転載)．

SIAM/ASA Journal on Uncertainty Quantification (JUQ) publishes research articles presenting significant mathematical, statistical, algorithmic, and application advances in uncertainty quantification, defined as the interface of complex modeling of processes and data, especially characterizations of the uncertainties inherent in the use of such models. The journal also focuses on related fields such as sensitivity analysis, model validation, model calibration, data assimilation, and code verification. The journal also solicits papers describing new ideas that could lead to significant progress in methodology

*takeru.matsuda@riken.jp
†miyatake@cas.cmc.osaka-u-u.ac.jp

for uncertainty quantification as well as review articles on particular aspects. The journal is dedicated to nurturing synergistic interactions between the mathematical, statistical, computational, and applications communities involved in uncertainty quantification and related areas. JUQ is jointly offered by SIAM and the American Statistical Association.

このように，数値解析と確率・統計のつながりは急速に重要性を増しており，両分野で蓄積されてきた方法論を融合・発展させる研究が今後ますます進展していくと考えられる．その中の一つに，ベイズ統計の観点から数値計算を捉え直す Probabilistic Numerics という分野がある．

1.2　ベイズ統計

ベイズ統計とは，ベイズの公式に基づいた統計学の枠組みで，機械学習やデータ同化でも多くの手法の基礎になっている [31]．いま，データ D が未知母数 θ をもつ確率分布 $p(D \mid \theta)$ に従う確率変数とする．ベイズの公式とは，事前分布（prior，θ に関する事前知識を確率分布で表現したもの）と尤度（likelihood，D から得られる θ に関する情報）を統合して事後分布（posterior，D をもとに更新した θ の確率分布）を構成する以下の計算式である：

$$\underbrace{\pi(\theta \mid D)}_{\text{posterior}} \propto \underbrace{\pi(\theta)}_{\text{prior}} \underbrace{p(D \mid \theta)}_{\text{likelihood}}.$$

適切に事前分布を設計することで，柔軟なデータ解析・モデリングを実現できるのがベイズ統計の大きな利点である．

ベイズ統計では未知量に関する知識が（点ではなく）分布の形で得られるため，自然に不確実性定量化が行えると期待される．たとえば，θ の事後分布が平均 μ，分散 σ^2 の正規分布として得られたとき，95 パーセントの確率で θ は区間 $[\mu-2\sigma, \mu+2\sigma]$ の中にあると考えられる．一般に，事後分布からこのように構成される区間のことを信用区間 (credible interval) という．信用区間を用いることで θ の不確実性が適切に定量化できるように思われるが，この方法は理論的には非自明な問題を含んでいる．すなわち，事後分布が事前分布に依存している以上，信用区間もまた事前分布に依存しており，客観的に事前分布をどう選択するかが明らかでない[*1]．信用区間の客観性を保証するために重要になるのが信頼区間 (confidence interval) の概念である．いま，データを入力して区間を出力する「手続き」を考えよう（上記の信用区間の構成もその一つ）．つまり，出力される区間はデータに依存してランダムに定まる．この手続きが「95 パーセント信頼区間を与える」とは，「母集団分布から生成されるデータを入力したとき，出力される区間が θ の真の値を含む確率」（被覆確率）が 95 パーセントになることである．信頼区間によって未知量を推定する問題を区間推定といい，仮説検定と双対の問題になっている [36]．信頼区間を用いることで未知母数に対する客観的な不確実性定量化が得られる．したがって，ベイズ統計によって客観的な不確実性定量化を行うには，信用区間が信頼区間の要件を満たしていなければならない．つまり，信用区間の被覆確率が（近似的に）95 パーセントになっている必要がある．この条件をみたす事前分布はマッチング事前分布と呼ばれる [10]．

ベイズ統計を用いたデータ解析を実行する際，（特に大規模複雑モデルでは）事後分布が解析的に表現できず計算機で近似的に取り扱う必要が生じることが多い．この問題に対して，マルコフ連鎖モンテカルロ (MCMC)，変分ベイズ，近似ベイズ計算 (ABC) など多くのアルゴリズムが開発されており，数値解析で培われてきた技術も活用されている．たとえば，マルコフ連鎖モンテカルロ (MCMC) はマルコフ連鎖を用いて事後分布からサンプリングを行う手法の総称であるが，MCMC の一手法であるハミルトンモンテカルロ (HMC) では常微分方程式に対するシンプレクティック解法の考え方を用いることでサンプリング効率を高めることに成功している [4]．ベイズ統計に限らず一般に統計解析に必要なアルゴリズムを開発する分野は計算統計 (Computational Statistics) と呼ばれ，「統計学のための数値解析」と捉えることができる．逆に，「数値解析のための統計学」ともいえるのが Probabilistic Numerics である．

[*1] よりマイナーな点としては，そもそも事後分布が単峰でないと信用区間を連結区間として定められないということもある．

1.3 Probabilistic Numerics

数値計算は，計算機の演算結果をもとに未知の量を推定する手続きと見なすことができる．この観点に立って既存の数値計算法をベイズ統計的に再解釈することで，アルゴリズムの改良や不確実性定量化への応用を目指すのが Probabilistic Numerics である [16, 7]．このようなアイデア自体は古くから存在したようだが [23, 11, 29]，近年再び脚光を浴びている．本節では，Probabilistic Numerics の最近の研究の一部をごく簡単に紹介する[*2]．

1.3.1 数値積分

分点 $x_1, \dots, x_N$ と重み $w_1, \dots, w_N$ を用いた区間 $[0, 1]$ 上の数値積分を考える：

$$I = \int_0^1 f(x)\mathrm{d}x \approx \sum_{i=1}^{N} w_i f(x_i).$$

これは，未知の積分値 I をデータ $(x_1, f(x_1)), \dots, (x_N, f(x_N))$ をもとに推定する問題と捉えることができる[*3]．ベイズ統計の枠組みで定式化するには被積分関数 f に事前分布を設定する必要があるが，関数空間上の確率分布（確率過程）として統計・機械学習でよく用いられるのがガウス過程である．ガウス過程は直感的には無限次元の正規分布であり，再生核ヒルベルト空間と密接な関わりをもつ．詳しくは [20] を参照されたい．ガウス過程を事前分布とすると積分値の事後分布が正規分布になり，事後平均に対応する重み $w_1, \dots, w_N$ も（多くの場合で）陽に得られる [5]．たとえば，被積分関数 f がブラウン運動に従うという事前分布を置いたとき，積分値 I の事後平均は台形則による計算結果と一致する [11]．この方法は多次元積分に対しても同様に適用できる．被積分関数 f がソボレフ空間に属するといった滑らかさに関する事前知識をガウス過程で表現したときの積分値の誤差の収束レートが詳細に解析されている [21]．このように数値積分をベイズ統計の枠組みで定式化する利点として，事後分布をもとに積分値の誤差の見積もりが得られることが挙げられる [5]．

1.3.2 微分方程式

常微分方程式の初期値問題を考える：

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = f(x(t)), \quad x(0) = x_0.$$

解 $x(t)$ を求めるために，ルンゲクッタ法をはじめとする多くの数値解法が存在する．この問題も，$x(t)$ の事前分布としてガウス過程を設定し，勾配 f の有限個の点での評価値をデータとみなすことでベイズ統計的に定式化できる．特に，時間の前後関係が存在するため，データ同化などで現れる逐次ベイズフィルタと似た設定になっている．この観点から，種々の離散化スキームとガウス過程の関係が判明してきている [33, 37, 34]．また，数値解の各ステップでの離散化誤差を確率変数とみなすアプローチもある [9, 1]．このように常微分方程式の離散化に伴う不確実性を確率によって定量化することで，自然に統計解析へ取り込めるという利点がある．たとえば，離散化に伴う不確実性を適切に尤度に反映させることで，未知パラメータに対する事後分布が誤った値に集中するのを抑えることができる [9]．初期値問題以外にも，境界値問題 [19] や偏微分方程式 [27] に対する手法も整備されてきている．

[*2] Probabilistic Numerics のウェブサイト（`http://probabilistic-numerics.org`）には関連文献や研究集会の情報がまとめられている．

[*3] 被積分関数は無限次元の自由度をもつが興味があるのは積分値（スカラー量）のみである．これはセミパラメトリック統計の問題設定に近いので，類似の手法が有効かもしれない．

1.3.3　線形計算

線形方程式 $Ax = b$ に対する代表的な反復解法である共役勾配法についてもベイズ統計的な解釈が得られている．．各反復で A^{-1} を逐次ベイズ推定しているとみなす立場 [15] と x を逐次ベイズ推定しているとみなす立場 [8] の 2 つが存在する（例えば [3] は両者の立場を整理したレビュー論文である）．後者では，x の事前分布を正規分布に設定すると共役性から事後分布も正規分布になり，その平均と分散共分散行列は陽に計算できる．よって，分散共分散行列に基づいて数値解の誤差（不確実性）の定量化が得られる*4．事後分布に基づいて良好な誤差評価が得られることが実験的に示されているが，その理論保証は今後の課題である．また，固有値・特異値の数値計算についてはほぼ未開拓のようである．

1.3.4　その他

非線形最適化についても，準ニュートン法 [14] やラインサーチ [24] に関するベイズ統計的解釈が得られている．前者では，準ニュートン法におけるヘッシアンの更新則を逐次ベイズ推定として定式化することで，より効率の良い更新則を提案している．

丸め誤差を確率変数とみなして区間演算に応用する試みもある [17]．通常の区間演算では答を「必ず」含む区間を出力するのに対して，ここでは答を「高確率で」含む区間を出力するという問題設定を考えている．これは最近さかんに研究されている高次元統計学でもよく現れる設定であり，集中不等式を用いる点も類似している．小さい確率で答を外すことを許す分，区間の幅を抑えることができる．

1.4　今後の展望

本節では，Probablistic Numerics に関連した今後の展望を述べる．

1.4.1　経験ベイズ信用区間の理論保証

ベイズ統計では事前分布の選択 (prior selection) が常に問題となる．事前分布の選択に関して，種々の視点から有効性が示されてきた手法が経験ベイズ法である [12]．これは，事前分布に含まれるハイパーパラメータ（例：ガウス過程の共分散関数のスケール・バンド幅）をデータから周辺尤度最大化によって定めるというものである．不確実性定量化の観点からは，経験ベイズ法によって構成した信用区間の被覆確率の理論保証は重要な課題である．この問題は従来のベイズ統計学でも長く研究されてきたが，Probablistic Numerics では問題設定がやや異なるため注意が必要である．たとえば，上述の数値積分や微分方程式に対する Probablistic Numerics の手法では未知の関数に対してガウス過程を事前分布に設定しているが，関数評価自体には確率的な誤差は仮定していない．一方，統計学で未知の関数を推定する問題であるノンパラメトリック回帰にもガウス過程は用いられるが，こちらでは関数値に確率的な誤差が加わったものがデータとして観測されると考えるのが一般的である．よって，後者の設定での経験ベイズ信用区間に対する理論保証 [13] は Probablistic Numerics の文脈にそのまま適用することはできず，研究が進行中である [22]．数値積分の確率的精度保証などへの応用が期待される．同様に，ベイズ共役勾配法 [8] において事前分布のスケールなどを経験ベイズ法で定めたときの理論保証も興味深い問題である．

そもそも，Probablistic Numerics では被覆確率をどの確率空間に対して定義するのかが自明でないと思われる．一つの考え方として，インスタンスに確率分布を想定するというものがある [30]．たとえば，数値積分における被積分関数に確率分布を仮定することが該当する．また，モンテカルロ法のように数値計算法自体が乱数を伴う場合は，その確率空間を対象とすることも考えられる．

---

*4ただし，事前分布の分散共分散行列によっては x の事後平均が共役勾配法の出力とは一致しないことに注意する．

1.4.2　構造保存数値解法・モデル縮減の確率的解釈

微分方程式を数値的に解く際，もとの方程式のもつ性質を数値解でも保存する「構造保存数値解法」が有効であることがよく知られている．Probabilistic Numerics の観点でこれを解釈すると，厳密解がもつ性質を事前分布の形に表現することで構造保存型の確率的数値解法を設計するという方針が考えられる．たとえば，線形保存量がある微分方程式に対して，解の事前分布を線形部分空間に制約すれば事後分布も同じ線形部分空間に制約されるため，保存量が数値解でも「保存」される[*5]．

大規模な微分方程式を低次元に縮約して計算量を削減する「モデル縮減」の発想を Probabilistic Numerics に応用することも考えられる．たとえば，代表的なモデル縮減手法の一つである Proper Orthogonal Decomposition (POD) では微分方程式を低次元線形部分空間に射影するが，これは統計学における基本的な次元削減手法である主成分分析に対応する．主成分分析は，データが低次元部分空間に乗っているという情報を事前分布に表現することでベイズ統計的に定式化することができる．同様の定式化によって POD を不確実性定量化に応用することが可能かもしれない．

以上のように，解の性質に関する事前知識を取り込んだ数値解法はベイズ統計と相性が良く，不確実性定量化への展開が期待される．

2　離散化誤差の定量化に基づく常微分方程式のパラメータ推定

本章では，常微分方程式の離散化に伴う不確実性を確率によって定量化するという発想を，常微分方程式の逆問題（観測データを用いたパラメータ推定）に利用した筆者らの最近の研究 [25] の概略を紹介する．

2.1　研究の動機（常微分方程式の逆問題）

Probabilistic Numerics 全般の研究背景は前述の通りだが，ここでは，常微分方程式の逆問題の視点から筆者らの研究の動機を述べる．

常微分方程式のパラメータ推定は，現代科学の様々な文脈において基本的な問題設定としてあらわれる．特に，初期値問題のパラメータ推定を行う場合，微分方程式の解が時系列観測データに何らかの意味でよく当てはまるように推定を行うことが多い．従って，長い時系列データを用いるほどより良い推定ができることが期待される．しかし，実用上は，微分方程式の厳密解の代わりに数値解をデータに当てはめて推定を行わざるをえないことも多い．離散化の影響により，数値解の誤差（= 厳密解と数値解の差）は時間発展とともに増大していくことが多く，パラメータを推定するにあたって，観測ノイズの影響よりも数値解の誤差の影響が支配的になる場合がある．このような場合，数値解を過度に信頼して推定を行うと，推定結果に大きなバイアスが生じ，長い時系列データを利用するほどかえって推定を悪化させる可能性もある．

この研究で主に想定している状況は，常微分方程式の性質や計算資源の制約などにより，それぞれの応用分野で要求される現実的な計算時間では十分に高精度な数値解が得られない（あるいはその数学的保証が困難な）場合である．このような場合に，パラメータ推定の文脈に適した形で，できるかぎり低コストで数値解の誤差を定量化できれば，より良い推定やより適切な信頼区間の構成に繋がることが期待される．

なお，以下で述べる内容においては，現代の数値解析の高度な知見を十分に活用できてはいない．しかし，これまでの数値解析や精度保証数値計算の文脈で培われてきた知見と反目するような内容ではなく，むしろ今後，分野融合的な研究に繋がることを強く予感させるものであると考えている．

2.2　問題設定

問題設定を述べる．次の常微分方程式の初期値問題を考えよう：

[*5] 非線形保存量がある微分方程式に対しても，解の事前分布を多様体に制約すれば事後分布も同じ多様体に制約される．ただし，一般に事後分布の平均は多様体からはみ出すことに注意する．

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t;\theta) = f(x(t;\theta)), \quad x(0;\theta) = \theta.$$

ただし，$f:\mathbb{R}\to\mathbb{R}$ であり，θ は推定したい初期値である．なお，本稿では以下で述べるアイデアの分かりやすさを優先して従属変数はスカラーとして議論を進めるが，ベクトルの場合も（以下の観測モデルに若干の追加の仮定が必要にはなるが）ほぼ同様に扱える．また，ベクトル場 f に未知パラメータが含まれる場合へも拡張できる．そこで本稿では，θ を初期値ではなくパラメータと呼ぶことにする．時刻 t_k $(0 \le t_1 < t_2 < \cdots < t_K)$ において，ノイズ付きの観測が得られているとしよう：

$$y_k = x_k(\theta) + \varepsilon_k, \quad \varepsilon_k \sim \mathrm{N}(0,\gamma^2), \quad k = 1,\ldots,K. \tag{1}$$

ここで，厳密解 $x(t_k;\theta)$ を $x_k(\theta)$ と表している．多くの応用では，観測ノイズの分散の推定も重要な課題だが，ここでは，分散は一旦既知と仮定する．

観測モデル (1) のもとで，θ を推定することを考えよう[*6]．特に，θ を最尤推定することを考えると，観測モデル (1) のもとで尤度関数が

$$L(\theta) = p(y_1,\ldots,y_K \mid \theta) = \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\gamma^2}} \exp\left(-\frac{(y_k - x_k(\theta))^2}{2\gamma^2}\right)$$

と書けることから，最尤推定量は

$$\hat{\theta}_{\mathrm{ML}} = \operatorname*{argmin}_{\theta\in\Theta} \sum_{k=1}^{K} \frac{(y_k - x_k(\theta))^2}{\gamma^2}$$

であるが，数値解を用いて推定を行う場合は

$$\hat{\theta}_{\mathrm{QML}} = \operatorname*{argmin}_{\theta\in\Theta} \sum_{k=1}^{K} \frac{(y_k - \tilde{x}_k(\theta))^2}{\gamma^2} \tag{2}$$

となる（$\tilde{x}_k(\theta) \approx x_k(\theta)$ は数値解を表す）．目的関数が異なるため一般に $\hat{\theta}_{\mathrm{ML}} \neq \hat{\theta}_{\mathrm{QML}}$ であり，特に数値解の誤差が無視できない場合には大きなバイアスが生じうる．

2.3 離散化誤差を考慮した推定手法

離散化による不確実性を確率変数によって定量化することで安定したパラメータ推定を行う手法について説明する．なお,「確率変数によって定量化する」というアイデアは既存研究でも採用されているものだが，具体的な定量化手法は既存研究とは異なるものとなっている．

数値解と厳密解の差を $\xi_k = \tilde{x}_k - x_k$ とおく．ここで，本来 ξ_k は一意に定まる値であるが，これを定量的に評価することを目指し，一旦，$\xi_1,\ldots,\xi_K$ は独立な正規分布に従う確率変数と仮定する：

$$\xi_k \sim \mathrm{N}(0,\sigma_k^2), \quad k = 1,\ldots,K. \tag{3}$$

平均ゼロの正規分布を考える妥当性は，微分方程式とその数値解法に応じて議論されるべきだが，ここでは最初の検討として平均ゼロの正規分布を考えている．分散 σ_k^2 は誤差の大きさを定量化する指標として導入したものであり，この段階ではその値は未知である．以後，この分散を「離散化誤差分散」と呼ぶ．離散化誤差のモデル (3) と観測モデル (1) を組み合わせると

$$y_k = \tilde{x}_k(\theta) + e_k, \quad e_k \sim \mathrm{N}(0,\gamma^2+\sigma_k^2), \quad k = 1,\ldots,K. \tag{4}$$

[*6]実際の応用では，観測だけではなく，モデル（微分方程式）そのものにも不確実性があることが多い [6] が，本稿ではモデルの不確実性は扱わない．

というモデルが得られる．このモデルをもとにパラメータ θ および離散化誤差分散 $\Sigma = \{\sigma_1^2, \ldots, \sigma_K^2\}$ を同時に推定することを考えよう．ここで，Σ は離散化誤差を定量化する指標として導入したが，パラメータ推定の観点では，数値解の信頼性を表す指標として解釈するのが自然であろう．そこで，離散化による誤差は時間発展とともに増大する（言い換えれば，数値解の信頼性は時間発展とともに低下する）ことが多いことを反映させるべく，Σ に単調増大性

$$0 \le \sigma_1^2 \le \sigma_2^2 \le \cdots \le \sigma_K^2 \tag{5}$$

を仮定する．単調増大性を満たす Σ の集合を $\mathcal{S}$ で表す．

すると，観測モデル (4) のもとで尤度関数は

$$L(\theta, \Sigma) = p(y_1, \ldots, y_K \mid \theta, \Sigma) = \sum_{k=1}^{L} \frac{1}{\sqrt{2\pi(\gamma^2 + \sigma_k^2)}} \exp\left(\frac{(y_k - \tilde{x}_k(\theta))^2}{2(\gamma^2 + \sigma_k^2)}\right) \tag{6}$$

となり，従って最尤推定量は

$$(\hat{\theta}, \hat{\Sigma}) = \operatorname*{argmin}_{(\theta,\Sigma)\in\Theta\times\mathcal{S}} \sum_{k=1}^{K} \left(\log\bigl(\gamma^2 + \sigma_k^2\bigr) + \frac{r_k^2(\theta)}{\gamma^2 + \sigma_k^2} \right), \quad r_k^2 = (y_k - \tilde{x}_k(\theta))^2 \tag{7}$$

と定式化される．

単調増大性 (5) について注意を述べる．離散化誤差分散 Σ について全く条件を課さずに尤度関数 (6) を最大化することを考えると，r_k^2 の増減に強く依存しすぎてしまい，かえって推定を悪化させてしまうことがある．そこで，微分方程式論や数値解析学の知見を利用して何らかの仮定をおく必要があり，ここではシンプルに単調増大性を考えている（単調性の仮定がどのような場合においても適切であると主張するものではない）．

2.4 最適化問題の解法

最適化問題 (7) の解法を考えよう．離散化誤差を考慮しない定式化 (2) と比較して，計算コストを著しく増加させることなく解く（あるいは，意味のある近似解を求める）ことが期待される．

表記を簡単にするため，$w_k = 1/(\gamma^2 + \sigma_k^2)$ とおき，$\mathcal{S}$ に対応する集合を $\mathcal{W}$ とする．すなわち，

$$\mathcal{W} = \left\{ (w_1, \ldots, y_K) \mid 0 < w_K \le \cdots \le w_1 \le \frac{1}{\gamma^2} \right\}$$

であり，w_k に関する不等式制約は，数値解の信頼性が時間発展とともに低下することを表している．このとき，最小化問題 (7) は

$$(\hat{\theta}, \hat{w}) = \operatorname*{argmin}_{(\theta,w)\in\Theta\times\mathcal{W}} g(\theta, w), \quad g(\theta, w) = \sum_{k=1}^{K} (-\log w_k + w_k r_k^2(\theta))$$

と表現できる．この問題を θ と w に関して交互に最小化する．すなわち，パラメータの適当な初期点 $\theta^{(0)}$ から出発し，

$$w^{(l)} = \operatorname*{argmin}_{w\in\mathcal{W}} g(\theta^{(l-1)}, w), \tag{8}$$

$$\theta^{(l)} = \operatorname*{argmin}_{\theta\in\Theta} g(\theta, w^{(l)}) \tag{9}$$

を収束するまで（あるいは，自然数 L を固定し $l = L$ まで）繰り返す．ここで，(9) は $w_k^{(l)}$ を重みとした重み付き最小二乗法と解釈でき，以後，全体のアルゴリズムを「繰り返し重み付き最小二乗法」とよぶ．

重み w の更新

重みの更新 (8) は凸最適化問題になっており，isotonic 回帰の文脈で知られている PAVA (pool adjacent violators algorithm) [2, 32] と呼ばれる組合せ的なアルゴリズムによって，厳密な最適解を $\mathrm{O}(K)$ の計算量で効率的に計算できる．

パラメータ θ の更新

パラメータ θ の更新式 (9) は，重み付き最小二乗問題である．具体的な問題にも依存するが，最適化問題としての性質は (2) とほぼ同じである．様々な数値解法が考えられるが，例えば準ニュートン法のような解法を利用するならば，目的関数の θ に関する勾配が必要になる．しかし，目的関数は数値解を通して θ に依存しているため，微分の情報をどのように得るかは自明ではない．例えば，データ同化などの文脈ではアジョイント法を用いて勾配を近似的に計算することが多い．実は近年の構造保存数値解法の文脈において，$\tilde{x}_k$ がルンゲクッタ法による数値解の場合，目的関数の勾配を効率よく厳密に計算できることが示されており [35]，このような手法を援用することもできる（本稿筆者らによる拡張もある [18, 26]）．

2.5　今後の展望

我々の研究に関して，その特徴を簡単に整理し，今後の展望を述べる．紙面の都合上，数値実験結果などは筆者らのプレプリント [25] に譲るが，Lorenz 方程式などで検証している範囲では，次のような特徴が観察されている．

- 提案手法は実際の誤差の振る舞いをある程度定量的に捉えることができ，また，離散化誤差を考慮しない手法に比べてパラメータ推定の精度も向上する傾向がある．
- 繰り返し重み付き最小二乗法は収束するまで繰り返す必要は必ずしもなく，2, 3 反復でも良好な結果が得られることが多い．
- 提案手法では観測ノイズの分散 γ^2 は既知と仮定しているが，γ^2 の事前知識として実際の分散よりも十分小さい値を用いた場合でも，最適化問題の解は大きく変わらない傾向があり，これは，分散の情報も同時に推定できる可能性を示唆している．

以上を踏まえ，本研究に関する今後の展望をいくつか述べる．まず，提案手法では確率変数が従う分布として正規分布を仮定するなど，様々な仮定がおかれているが，微分方程式や数値解法のタイプに応じて，仮定の妥当性や，あるいは他の仮定の検討などを行う必要がある．また，現状，最尤推定の枠組みで研究を行っているが，ベイズ推定などへの展開も興味深い．さらに，将来的に応用諸分野への展開を見据え，実問題などのより現実的な問題に対する検証なども行っていく必要がある．

参考文献

[1] Abdulle, A. & Garegnani, G. (2020). Random time step probabilistic methods for uncertainty quantification in chaotic and geometric numerical integration. *Statistics and Computing*, accpeted. (arXiv:1801.01340)

[2] Barlow, R. E., Bartholomew, D. J., Bremner, J. M. & Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression*. Wiley.

[3] Bartles, S., Cockayne, J., Ipsen, I. C. F. & Hennig, P. (2019). Probabilistic linear solvers: a unifying view. *Statistics and Computing* **29**, 1249–1263.

[4] Bou-Rabee, N. & Sanz-Serna, J. M. (2018). Geometric integrators and the Hamiltonian Monte Carlo method. *Acta Numerica*, **27**, 113–206.

[5] Briol, F-X., Oates, C. J., Girolami, M., Osborne, M. A. & Sejdinovic, D. (2019). Probabilistic integration: a role in statistical computation?. *Statistical Science* **34**, 1–22.

[6] Brynjarsdóttir, J. & O'Hagan, A. (2014). Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, **30**, 114007.

[7] Cockayne, J., Oates, C. J., Sullivan, T. & Girolami, M. (2019). Bayesian probabilistic numerical methods. *SIAM Review* **61**, 756–789.

[8] Cockayne, J., Oates, C. J., Ipsen, I. C. F. & Girolami, M. (2019). A Bayesian Conjugate Gradient Method. *Bayesian Analysis* **14**, 937–1012.

[9] Conrad, P. R., Girolami, M., Särkkä, S., Stuart, A. & Zygalakis, K. (2017). Statistical analysis of differential equations: introducing probability measures on numerical solutions. *Statistics and Computing* **27**, 1065–1082.

[10] Datta, G. S. & Mukerjee, R. (2004). *Probability Matching Priors: Higher Order Asymptotics*. Springer.

[11] Diaconis, P. (1988). Bayesian numerical analysis. Statistical Decision Theory and Related Topics IV, 1, 163–175.

[12] Efron, B. & Hastie, T. (2016). *Computer Age Statistical Inference*. Cambridge University Press.

[13] Ghosal, S. & van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press.

[14] Hennig, P. & Kiefel, M. (2013). Quasi-newton method: A new direction. *Journal of Machine Learning Research* **14**, 843–865.

[15] Hennig, P. (2015). Probabilistic interpretation of linear solvers. *SIAM Journal on Optimization* **25**, 234–260.

[16] Hennig, P., Osborne, M. A. & Girolami, M. (2015). Probabilistic numerics and uncertainty in computations. *Proceedings of the Royal Society of London. Series A* **471**, 20150142.

[17] Higham, N. J. & Mary, T. (2019). A new approach to probabilistic rounding error analysis. *SIAM Journal on Scientific Computing* **41**, 2815–2835.

[18] Ito, S., Matsuda, T. & Miyatake, Y. (2019). Adjoint-based exact Hessian-vector multiplication using symplectic Runge–Kutta methods. arXiv:1910.06524.

[19] John, D., Heuveline, V. & Schober, M. (2019). GOODE: A Gaussian Off-The-Shelf Ordinary Differential Equation Solver. International Conference on Machine Learning (ICML), 3152-3162.

[20] Kanagawa, M., Hennig, P., Sejdinovic, D. & Sriperumbudur, B. K. (2018). Gaussian Processes and Kernel Methods: A Review on Connections and Equivalences. arXiv:1807.02582.

[21] Kanagawa, M., Sriperumbudur, B. K. & Fukumizu, K. (2020). Convergence Analysis of Deterministic Kernel-Based Quadrature Rules in Misspecified Settings. *Foundations of Computational Mathematics* **20**, 155–194.

[22] Karvonen, T., Wynne, G., Tronarp, F., Oates, C. & Sarkka, S. (2020). Maximum likelihood estimation and uncertainty quantification for Gaussian process approximation of deterministic functions. arXiv:2001.10965.

[23] Larkin, F. M. (1972). Gaussian measure in Hilbert space and applications in numerical analysis. *Rocky Mountain Journal of Mathematics* **2**, 379–422.

[24] Mahsereci, M. & Hennig, P. (2017). Probabilistic line searches for stochastic optimization. *Journal of Machine Learning Research* **18**, 4262–4320.

[25] Matsuda, T. & Miyatake, Y. (2019). Estimation of ordinary differential equation models with discretization error quantification. arXiv:1907.10565.

[26] Matsuda, T. & Miyatake, Y. (2020). Generalization of partitioned Runge–Kutta methods for adjoint systems. arXiv:2003.09789.

[27] Oates, C. J., Cockayne, J., Aykroyd, R. G. & Girolami, M. (2019). Bayesian Probabilistic Numerical Methods in Time-Dependent State Estimation for Industrial Hydrocyclone Equipment. *Journal of the American Statistical Association* **114**, 1518–1531.

[28] Oates, C. J. & Sullivan, T. J. (2019). A modern retrospective on probabilistic numerics. *Statistics and Computing* **29**, 1335–1551.

[29] O'Hagan, A. (1992). Some Bayesian numerical analysis. Bayesian Statistics, 4, 345–363.

[30] Ritter, K. (2000). *Average-Case Analysis of Numerical Problems*. Springer.

[31] Robert, C. (2001). *The Bayesian Choice*. Springer.

[32] Robertson, T., Wright, F. T. & Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. Wiley.

[33] Schober, M., Duvenaud, D. K. & Hennig, P. (2014). Probabilistic ODE solvers with Runge–Kutta means. Advances in Neural Information Processing Systems (NIPS), 739–747.

[34] Schober, M., Sarkka, S. & Hennig, P. (2019). A probabilistic model for the numerical solution of initial value problems. *Statistics and Computing* **29**, 99–122.

[35] Sanz-Serna, J. M. (2016). Symplectic Runge–Kutta schemes for adjoint equations, automatic differentiation, optimal control, and more. *SIAM Review* **58**, 3–33.

[36] 竹村 彰通 (1991). 現代数理統計学. 創文社.

[37] Teymur, O., Zygalakis, K. & Calderhead, B. (2016). Probabilistic linear multistep methods. Advances in Neural Information Processing Systems (NIPS), 4321–4328.