Data assimilation for massive autonomous systems based on a second-order adjoint method

Hiromichi NAGAO (ERI/UTokyo)

S. Ito, A. Yamanaka, Y. Tsukada, T. Koyama, M. Kano, and J. Inoue

DA Seminar @ RIKEN AICS

Feb. 7, 2017

CV (Hiromichi NAGAO, 長尾大道)

1991-1995 Faculty of Science, Kyoto Univ.

1995-2002 Dept. of Geophysics, Graduate School of Science, Kyoto Univ.
(Doctorate of Science)

2002-2006 Researcher, Tono Geoscience Center,
Japan Nuclear Cycle Development Institute

2006-2009 Researcher, IFREE, JAMSTEC

2009-2013 Project Associate Professor, Research and Development Center for Data Assimilation, The Institute of Statistical Mathematics

2013-present

Associate Professor

Earthquake Research Institute, Univ. of Tokyo

Dept. of Mathematical Informatics,
Graduate School of Science and Technology, Univ. of Tokyo

The Institute of Statistical Mathematics

Outline

1. Data assimilation for structural materials
2. Adjoint method (4D-Var)
3. Application of the adjoint method to a phase-field model
4. Uncertainty quantification of estimates based on a second-order adjoint method
5. Validation of the proposed method through twin experiments
6. Summary

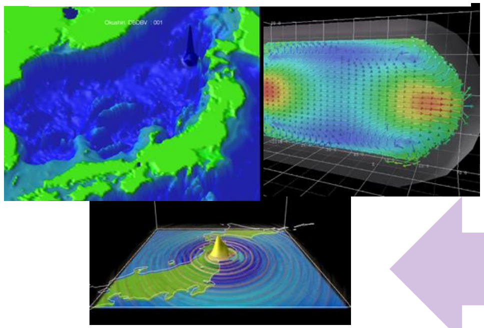
Today's talk is based on the following paper:

Ito, S., H. Nagao, A. Yamanaka, Y. Tsukada, T. Koyama, M. Kano, and J. Inoue, Data assimilation for massive autonomous systems based on a second-order adjoint method, Phys. Rev. E, 94, 043307, doi:10.1103/PhysRevE.94.043307, 2016.

Data Assimilation (DA)

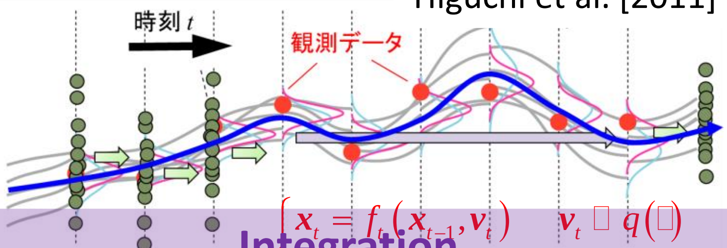
Integration of numerical simulations and observational data based on Bayesian statistics

Numerical Simulations



Bayesian Statistics, State Space Model

Higuchi et al. [2011]

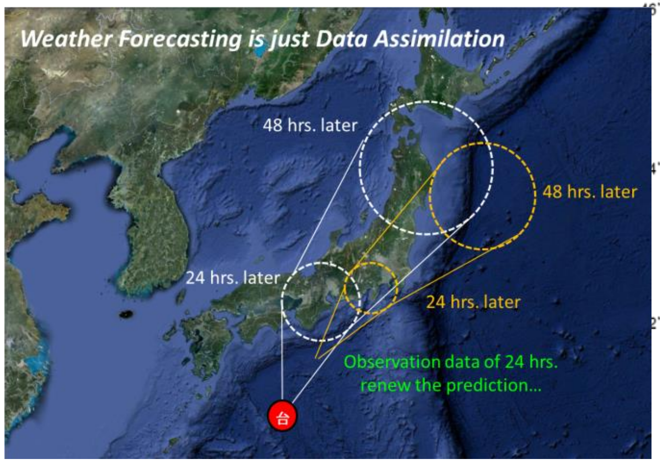


$$\begin{cases} x_t = f_t(x_{t-1}, v_t) & v_t \sim q(\cdot) \\ y_t = h_t(x_t, w_t) & w_t \sim r(\cdot) \end{cases}$$

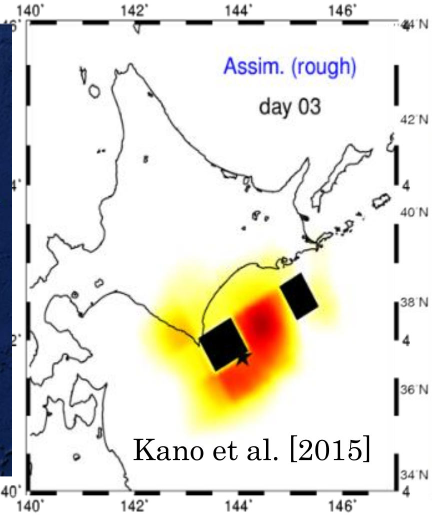
Integration

Observation Data

Weather Forecasting

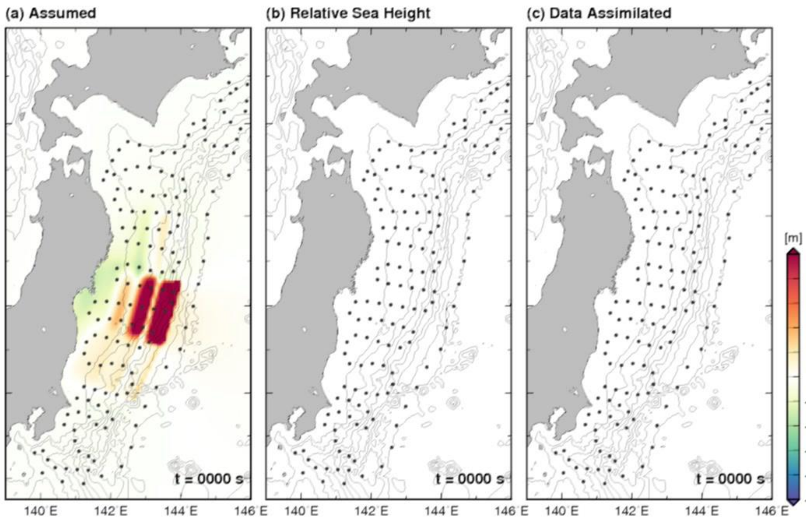


Afterslips



Tsunami

Maeda et al. [2015]

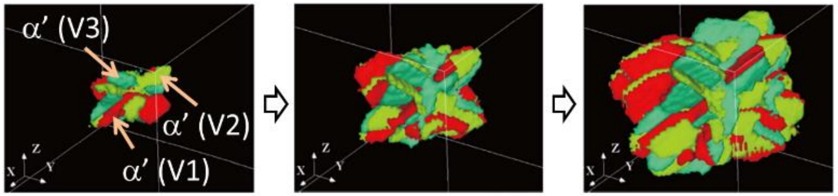


Development of Materials Integration System (SIP-MI)

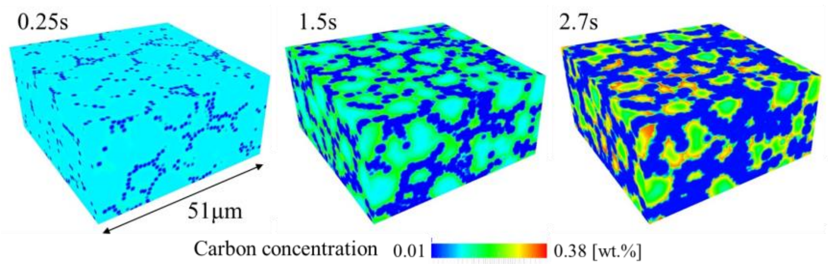
DA for Structural Materials

Estimation of parameters and internal states in materials

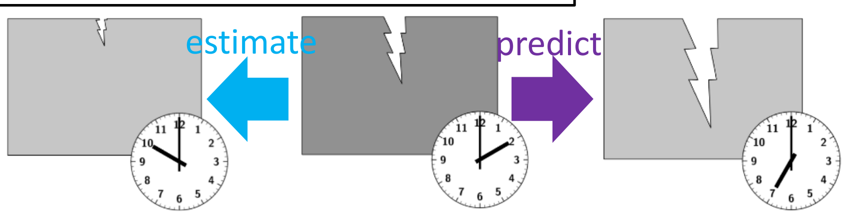
Simulation of martensitic transformation (Prof. Koyama)



Transformation from austenite to ferrite (Prof. Yamanaka)



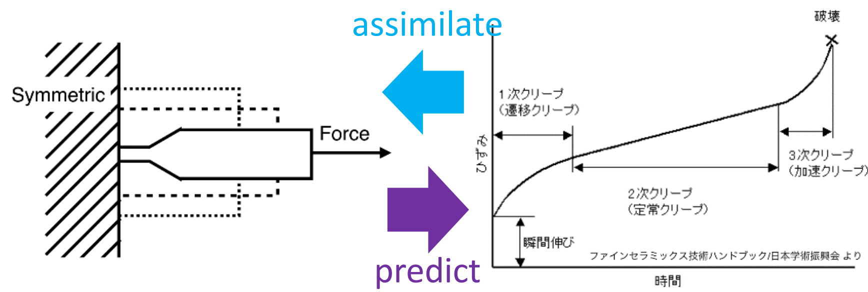
Forward / Inversion problems of cracks

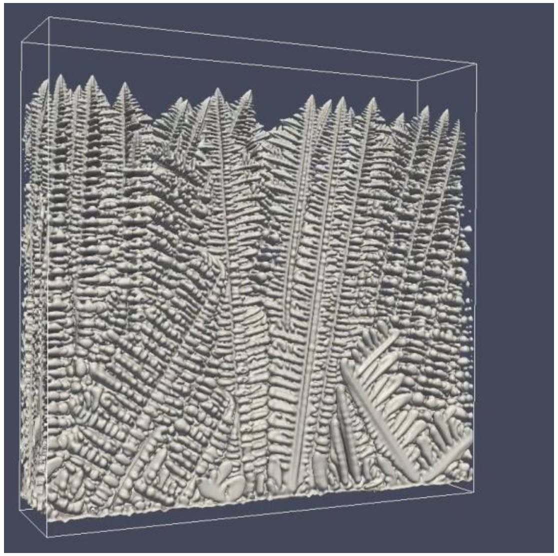


Development of DA beneficial to the phase-field method

➡ Parameter estimation, Experimental design, etc.

Estimation of creep rupture via elastic-plastic dynamics





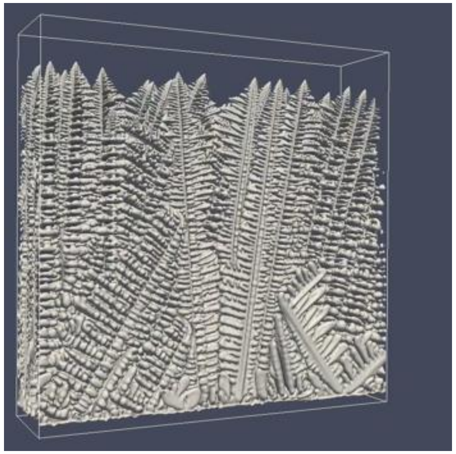
Shimokawabe et al. [2011]

DA for Systems Having Large Degrees of Freedom

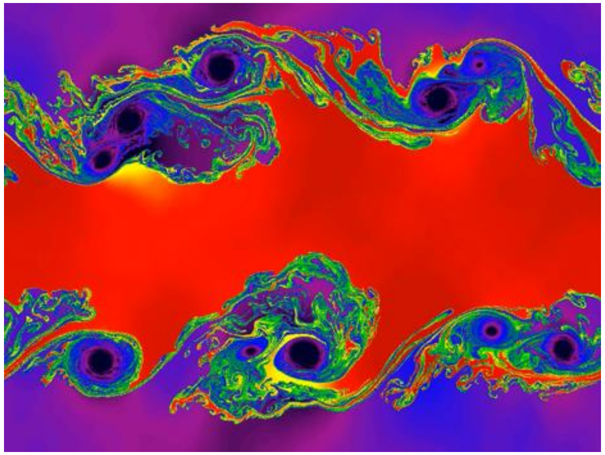
Numerical simulations in continuous fields

⇩

Massive computations



Phase-field model
(dendrite growth)
Shimokawabe et al. [2011]



Navier-Stokes equation
(K-H instability)
Springel [2009]

Sequential DA based on such as Kalman filter or particle filter

➡ requires memory of $O(N^2)$ (N: the degree of freedom)

e.g.
$N \sim 10^9 \Rightarrow \quad \sim 10^4$ Pbytes
cf. K computer ~ 1 Pbytes

DA method applicable to systems having large degrees of freedom is needed

State & Parameter Estimation based on Adjoint Method

System model

$$\frac{\partial \boldsymbol{\theta}_t}{\partial t} = \boldsymbol{F}(\boldsymbol{\theta}_t)$$

Observation model

$$D_t = h(\boldsymbol{\theta}_t) + \omega$$

Search $\boldsymbol{\Theta} = \boldsymbol{\theta}_0$ that best matches observation data

Bayes' theorem

$$p(\boldsymbol{\Theta}|\mathbf{D}) \propto p(\boldsymbol{\Theta})p(\mathbf{D}|\boldsymbol{\Theta})$$

posterior — prior — likelihood

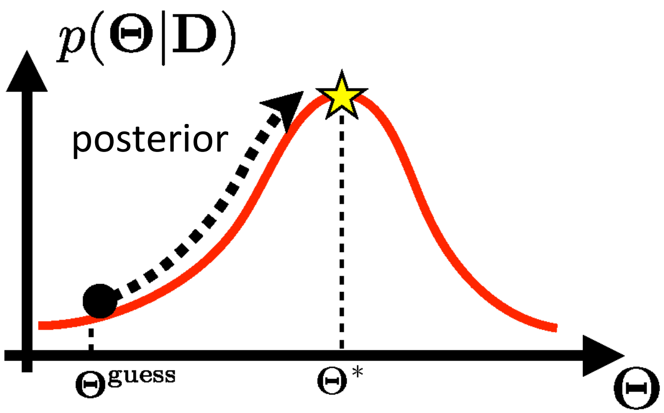
Cost function

$$J = -\log p(\boldsymbol{\Theta}) - \sum_{t \in \mathcal{T}} \log q\left(D_t - h(\boldsymbol{\theta}_t)\right)$$

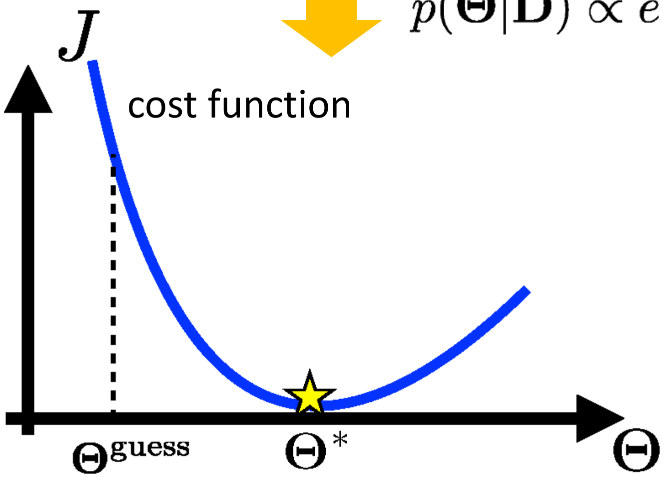
Adjoint equation

$$\frac{\partial \boldsymbol{\lambda}_t}{\partial t} + \left(\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\theta}_t}\right)^{\top} \boldsymbol{\lambda}_t = \frac{\partial \mathcal{J}}{\partial \boldsymbol{\theta}_t}$$

$$\boldsymbol{\lambda}_T = 0 \qquad \boldsymbol{\lambda}_0 = -\frac{\partial J}{\partial \boldsymbol{\Theta}}$$

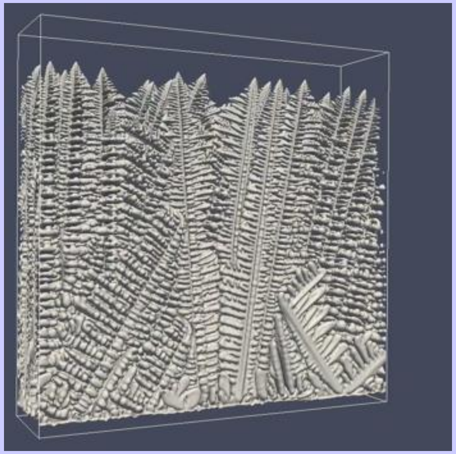


$$p(\boldsymbol{\Theta}|\mathbf{D}) \propto e^{-J}$$



State & Parameter Estimation based on Adjoint Method

Phase-field model



Needs massive computational costs due to fine grids

Sequential Bayesian filters

Computational time $e^{O(N)}$ (N: degree of freedom)

Searches randomly parameter space → Can estimate optimum and its uncertainty

Adjoint method (4D-Var)

Computational time $O(N)$

Estimates only optimum that maximizes posterior → Cannot evaluate uncertainty of optimum

We develop a DA method that can estimate not only optimum but also its uncertainty even in the case of a system having large degrees of freedom

Testbed: Phase-Field Model

Phase-field model describing interface migration Kobayashi [1993]

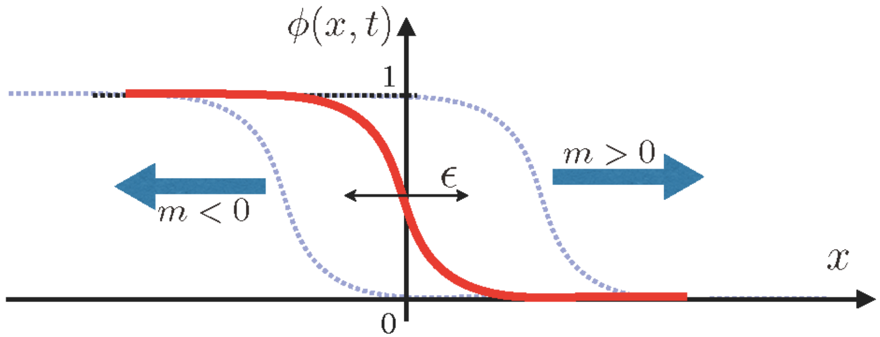
$$\tau \frac{\partial \phi}{\partial t} = \epsilon^2 \triangle \phi + \phi (1 - \phi) \left(\phi - \frac{1}{2} + m \right) \qquad |m| < \frac{1}{2}$$

$\tau, \ \epsilon, \ m$ are assumed to be constants in time and space

1D

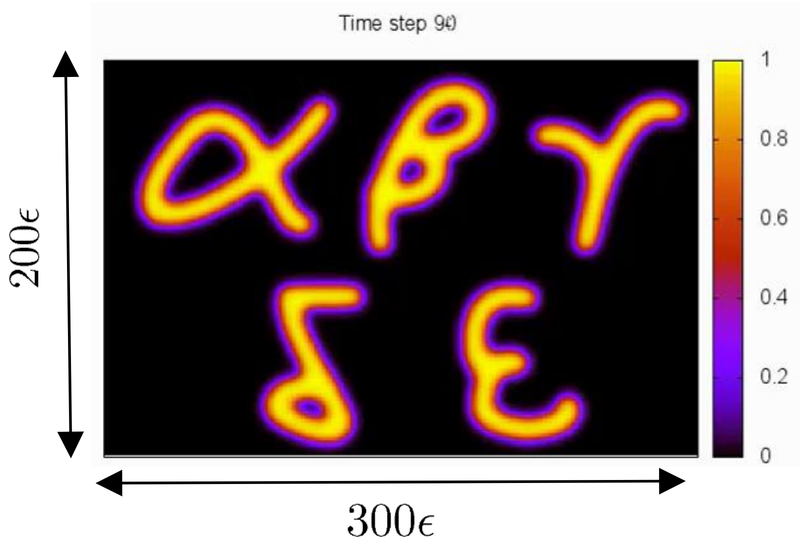
if $\begin{cases} m = \text{const.} \\ \phi(-\infty, t) = 1, \ \phi(\infty, t) = 0 \end{cases}$

$$\phi(x,t) = \frac{1}{2}\left[1 - \tanh\left(\frac{x}{2\sqrt{2}\varepsilon} - \frac{mt}{2\tau}\right)\right]$$



2D

$\Delta t = 0.1\tau \quad \Delta x = \epsilon$



First-Order Adjoint of Phase-Field Model

Forward

$$\tau \frac{\partial \phi_i}{\partial t} = \epsilon^2 \triangle_i \phi_i + \phi_i (1 - \phi_i) \left(\phi_i + m - \frac{1}{2} \right)$$

$$\frac{\partial m}{\partial t} = 0$$

Backward

$$-\tau \frac{\partial \tilde{\phi}_i}{\partial t} = \epsilon^2 \triangle_i \tilde{\phi}_i + \left\{ -3\phi_i^2 + (3 - 2m)\phi_i + m - \frac{1}{2} \right\} \tilde{\phi}_i - \frac{\partial \mathcal{J}}{\partial \phi_i} \sum_{k \in \mathcal{K}} \delta(t - t_k)$$

$$-\tau \frac{\partial \tilde{m}}{\partial t} = \sum_j \phi_j (1 - \phi_j) \tilde{\phi}_j - \frac{\partial \mathcal{J}}{\partial m} \sum_{k \in \mathcal{K}} \delta(t - t_k)$$

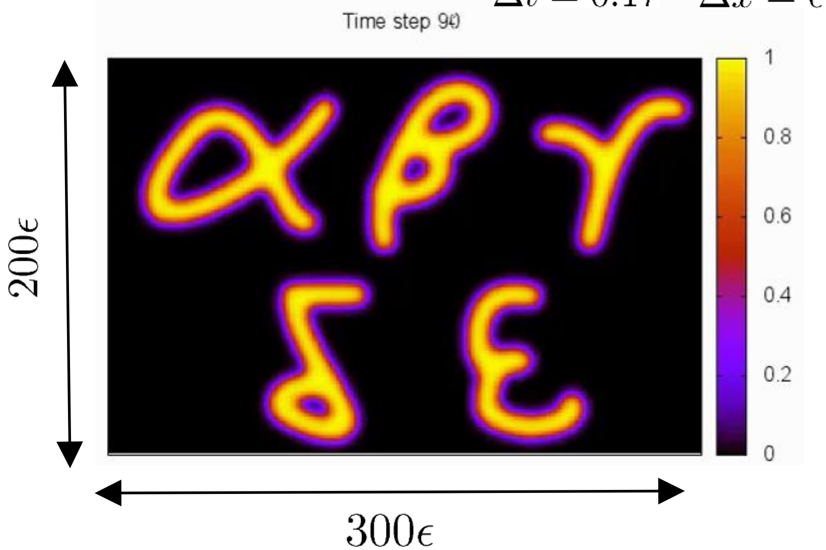
Twin Experiment

Suppose that we have observation data of the phase $\phi(\boldsymbol{x},t)$ in 2D, which satisfies

$$\tau \frac{\partial \phi}{\partial t} = \epsilon^2 \triangle \phi + \phi(1-\phi)\left(\phi - \frac{1}{2} + m\right) \qquad |m| < \frac{1}{2}$$

Estimate the parameter m and initial state $\phi(\boldsymbol{x},0)$ from the observation data contaminated by Gaussian noise, i.e.,

$\Delta t = 0.1\tau \quad \Delta x = \epsilon$



+ noise that follows $N\left(\mathbf{0},\sigma^2 I\right)$

System Model & State Variable

System model

$$\begin{cases} \dfrac{\partial \phi}{\partial t} = \dfrac{\epsilon^2}{\tau}\Delta\phi + \dfrac{1}{\tau}\phi(1-\phi)\left(\phi - \dfrac{1}{2} + m\right) \\ \dfrac{\partial m}{\partial t} = 0 \end{cases}$$

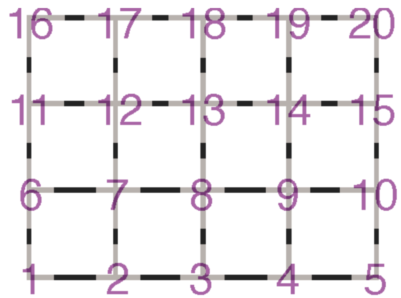
s.t. $|m| < \frac{1}{2}$ $\quad 0 < \phi(\boldsymbol{x}, 0) < 1$

State variable

$$\boldsymbol{\theta}(t) = \left(\boldsymbol{\phi}^\top, m + \frac{1}{2}\right)^\top$$

M: number of computational grids in space

$$\boldsymbol{\phi}(t) = (\phi_1, \phi_2, \cdots, \phi_M)^\top$$



$$\frac{\partial \boldsymbol{\theta}}{\partial t} = \boldsymbol{F}(\boldsymbol{\theta}) \quad \text{s.t.} \quad 0 < \theta_i(0) < 1$$

What we to do is to find $\boldsymbol{\theta}(0) = \boldsymbol{\Theta}$
that best matches observation data

Bayes' Theorem

state data

$$p(\boldsymbol{\Theta} | D) = \frac{1}{p(D)} p(\boldsymbol{\Theta}) p(D | \boldsymbol{\Theta})$$

constant

Posterior $\propto$ Prior $\times$ Likelihood

a priori information

misfit between model and data

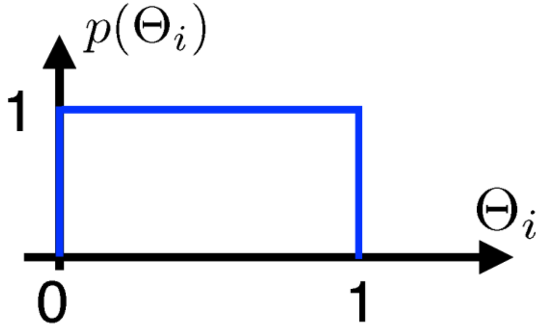
Prior

Bayes' theorem

$$p(\boldsymbol{\Theta}|D) = \frac{1}{p(D)} p(\boldsymbol{\Theta}) p(D|\boldsymbol{\Theta})$$

Prior

$$0 < \Theta_i < 1$$



$$p(\Theta_i) = \begin{cases} 1 & 0 < \Theta_i < 1 \\ 0 & \text{otherwise} \end{cases}$$

Likelihood

Bayes' theorem

$$p(\boldsymbol{\Theta}|D) = \frac{1}{p(D)} p(\boldsymbol{\Theta}) p(D|\boldsymbol{\Theta})$$

Likelihood

$$D_t = h(\boldsymbol{\theta}_t) + \omega \qquad \omega \overset{i.i.d}{\sim} N\left(0, \sigma^2\right)$$

$$p\left(D \,|\, \boldsymbol{\Theta}\right) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left[-\sum_{t \in T_{\mathrm{obs}}} \frac{\left(h\left(\boldsymbol{\theta}_t\right) - D_t\right)^2}{2\sigma^2}\right]$$

Optimization

$$\min_{\boldsymbol{\Theta}} p(\boldsymbol{\Theta} \mid D) \quad \Leftrightarrow \quad \max_{\boldsymbol{\Theta}} \{ \underbrace{-\log p(\boldsymbol{\Theta} \mid D)}_{J} \}$$

Cost function

$$J = \text{const.} + \frac{n}{2}\log 2\pi\sigma^2 + \frac{1}{2\sigma^2}\int dt \sum_{t_s \in T_{\text{obs}}} \delta(t - t_s)\left[h(\boldsymbol{\theta}(t_s)) - D(t_s)\right]^2$$

where $0 < \Theta_i < 1$

$$\boldsymbol{\Theta}^* = \arg\min_{\boldsymbol{\Theta}} J \quad \text{s.t.} \quad 0 < \Theta_i < 1$$

$\boldsymbol{\Theta}$ is to be optimized by a gradient method, but $\partial J / \partial \boldsymbol{\Theta}$ cannot be easily obtained since J does not include $\boldsymbol{\Theta}$ explicitly.

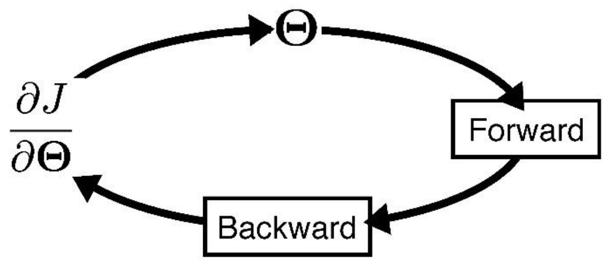
Adjoint Method (4D-Var)

$$J\left(\mathbf{\Theta}\right) = \int_0^T dt\, \mathcal{J}(\boldsymbol{\theta}(t)) + \int_0^T dt\, \boldsymbol{\lambda}^\top \cdot \left(\frac{\partial \boldsymbol{\theta}}{\partial t} - \boldsymbol{F}(\boldsymbol{\theta}) \right)$$

where $\mathcal{J}(\boldsymbol{\theta}(t)) = \sum_{t_s \in \mathcal{T}_{\mathrm{obs}}} \delta\left(t - t_s\right) \left[h\left(\boldsymbol{\theta}(t)\right) - D \right]^2$

Forward model

$$\frac{\partial \boldsymbol{\theta}}{\partial t} = \boldsymbol{F}(\boldsymbol{\theta}) \qquad \boldsymbol{\theta}(0) = \mathbf{\Theta}$$



Backward: Adjoint model

$$\frac{\partial \boldsymbol{\lambda}}{\partial t} + \left(\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\theta}} \right)^\top \cdot \boldsymbol{\lambda} = \frac{\partial \mathcal{J}}{\partial \boldsymbol{\theta}} \qquad \boldsymbol{\lambda}(0) = -\frac{\partial J}{\partial \mathbf{\Theta}}$$

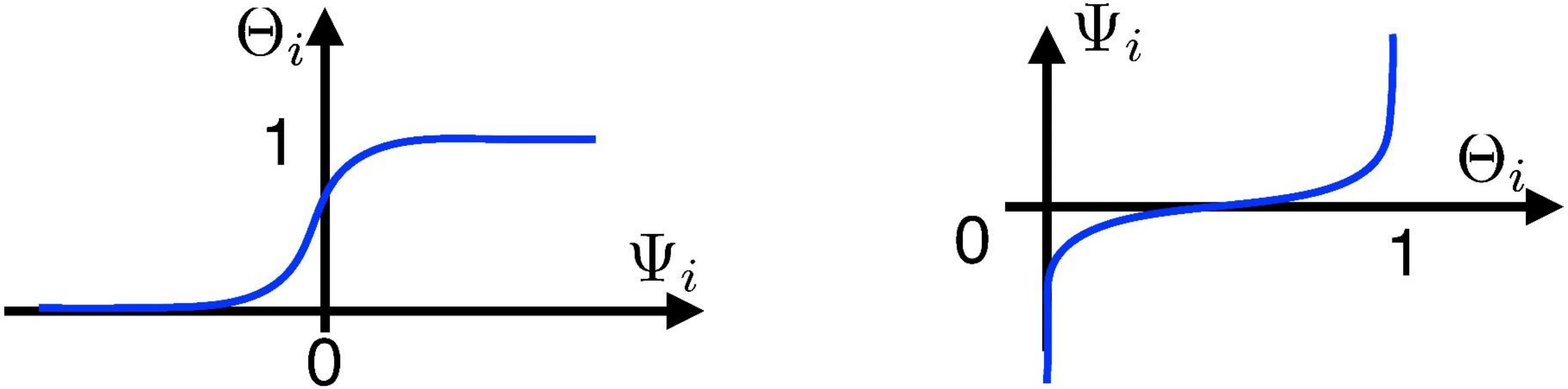
$$\boldsymbol{\lambda}(T) = 0$$

Variable Transformation for Constraint Condition

$$\Theta_i = \frac{1}{1+\exp(-\Psi_i)} \quad \Longleftrightarrow \quad \Psi_i = \log\left(\frac{\Theta_i}{1-\Theta_i}\right)$$

$$0 < \Theta_i < 1 \qquad -\infty < \Psi_i < \infty$$

$$\frac{\partial J}{\partial \Psi_i} = \Theta_i \left(1-\Theta_i\right) \frac{\partial J}{\partial \Theta_i}$$



Procedure of Adjoint Method

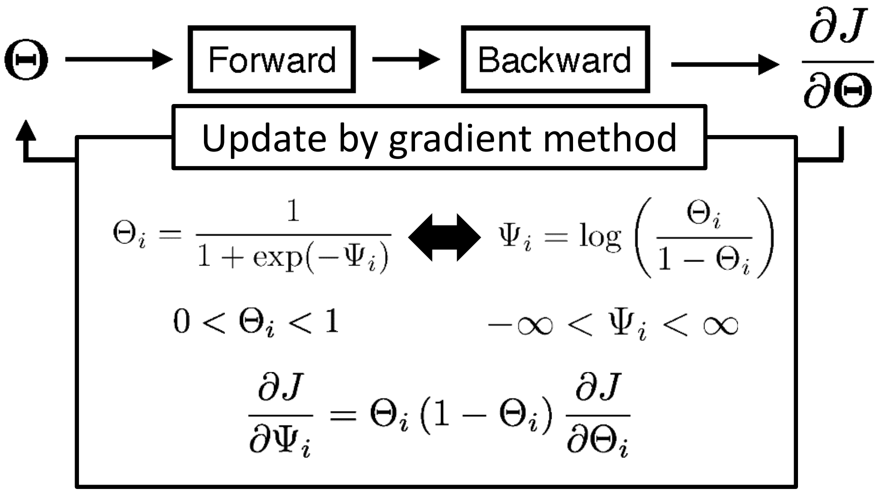
1. Give an initial value
2. Compute $\frac{\partial J}{\partial \boldsymbol{\Theta}}$ by an adjoint method
3. Transform $\boldsymbol{\Theta}$ to $\boldsymbol{\Psi}$ by

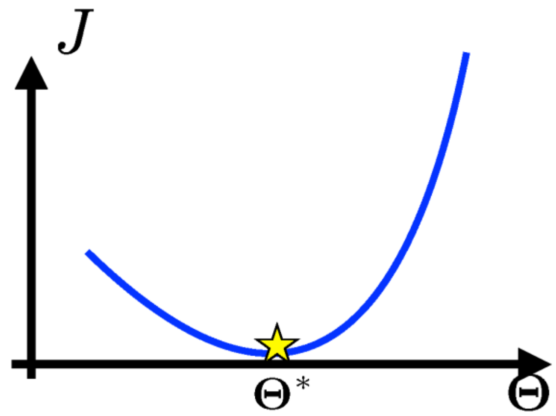
$$\Psi_i = \log\left(\frac{\Theta_i}{1-\Theta_i}\right) \qquad \frac{\partial J}{\partial \Psi_i} = \Theta_i\left(1-\Theta_i\right)\frac{\partial J}{\partial \Theta_i}$$

4. Update $\boldsymbol{\Psi}$ by a gradient method
5. Transform inversely $\boldsymbol{\Psi}$ to $\boldsymbol{\Theta}$ by

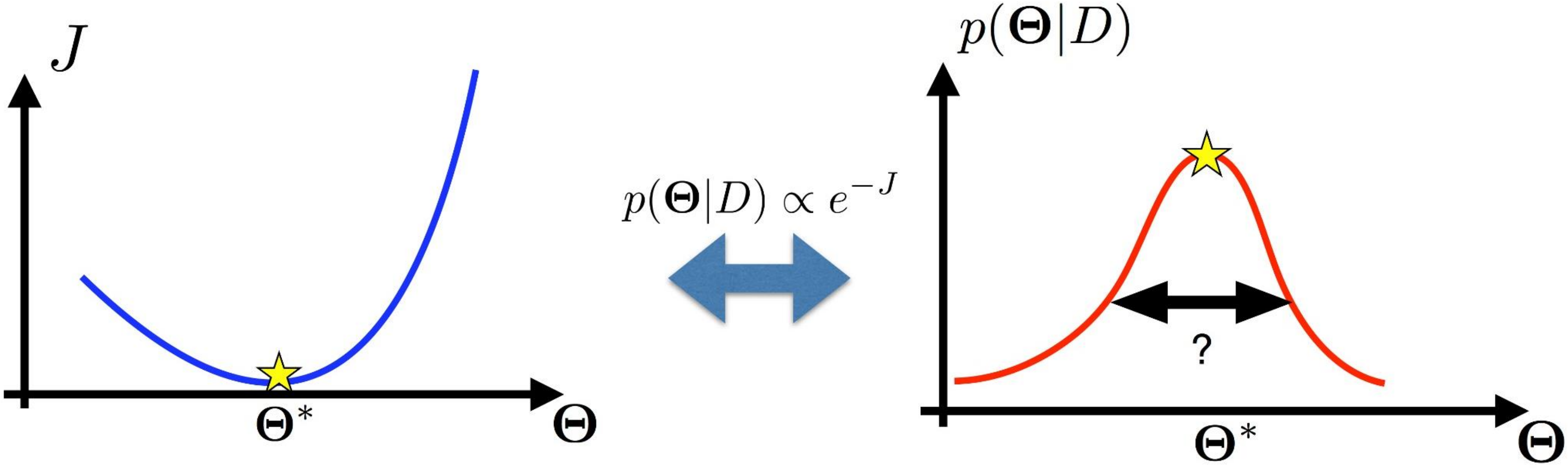
$$\Psi_i = \log\left(\frac{\Theta_i}{1-\Theta_i}\right)$$

6. Repeat 2.-5. until convergence





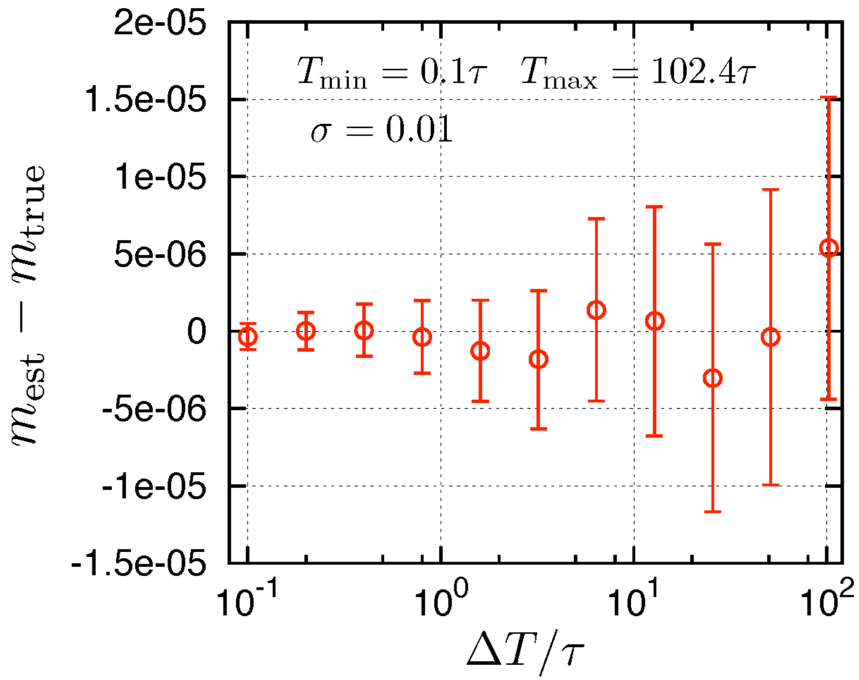
Problem in Adjoint Method



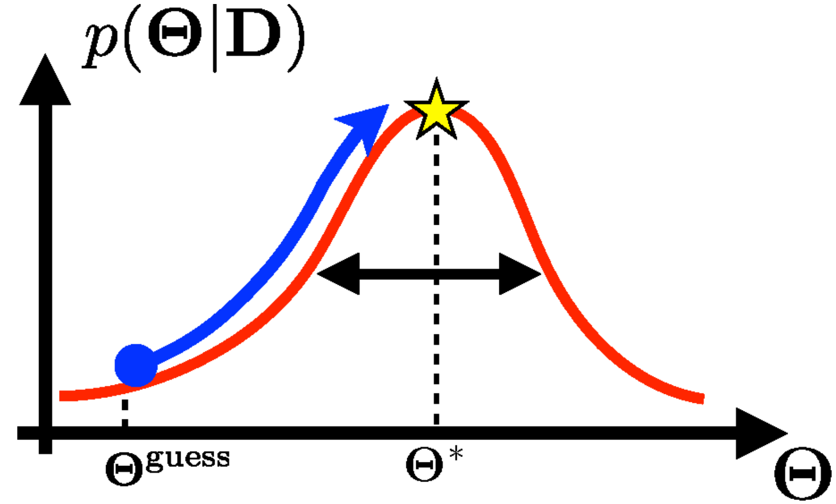
The current framework of adjoint method never evaluates the uncertainty of estimates

Uncertainty Quantification (UQ)

We have established a methodology of uncertainty quantification using second-order adjoint method



Gives feedback to experimental design!



Enables us to estimate optimum and evaluate its uncertainty

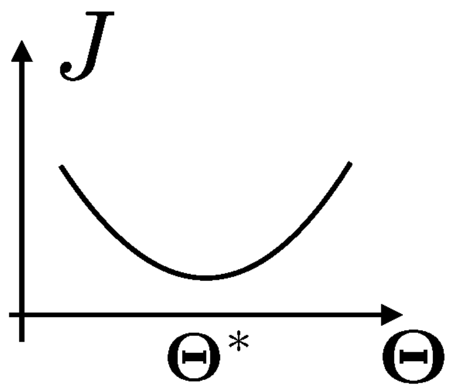
Ito, S., H. Nagao, A. Yamanaka, Y. Tsukada, T. Koyama, M. Kano, and J. Inoue, Data assimilation for massive autonomous systems based on a second-order adjoint method, Phys. Rev. E, 94, 043307, doi:10.1103/PhysRevE.94.043307, 2016.

Laplace Approximation of Posterior

Laplace approximation

Cost function can be approximated as a second-order polynomial in the neighborhood of the optimum $\boldsymbol{\Theta}^*$

$$J(\boldsymbol{\Theta}) \sim J(\boldsymbol{\Theta}^*) + \frac{1}{2}\left(\boldsymbol{\Theta} - \boldsymbol{\Theta}^*\right)^\top H \left(\boldsymbol{\Theta} - \boldsymbol{\Theta}^*\right)$$



$$p(\boldsymbol{\Theta}|D) \sim N(\boldsymbol{\Theta}^*, H^{-1})$$

H^{-1} : inverse of the Hessian matrix $H = \left.\frac{\partial^2 J}{\partial \boldsymbol{\Theta}^2}\right|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^*}$

Direct computation of H^{-1} requires unpractical computation time $O(N^3)$.

But, what we need are only a limited number diagonal elements of H^{-1}

$$p(\Theta_k|D) = \int d\Theta_{-k}\, p(\boldsymbol{\Theta}|D) = N\left(\Theta_k^*, (H^{-1})_{k,k}\right)$$

Second-Order Adjoint Method

We want to obtain only the k-th diagonal element in H^{-1} without explicitly computing H^{-1}

Solve $H\boldsymbol{r} = \boldsymbol{b}$ using an iterative method, where $\boldsymbol{b} = (0,\cdots,0,1,0,\cdots,0)^{\mathrm{T}}$

Needs a method to compute Hessian-vector product $H\boldsymbol{\alpha}$

Second-order adjoint method

Forward: Tangent linear model

$$\frac{\partial \boldsymbol{\xi}}{\partial t} = \left(\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\theta}}\right) \cdot \boldsymbol{\xi} \qquad \boldsymbol{\xi}(0) = \underline{\boldsymbol{r}}$$

input

Backward: 2nd-order adjoint model

$$\frac{\partial \boldsymbol{\zeta}}{\partial t} + \left(\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\theta}}\right)^{\top} \cdot \boldsymbol{\zeta} = \left(\frac{\partial^2 \boldsymbol{F}}{\partial \boldsymbol{\theta}^2} \cdot \boldsymbol{\xi}\right)^{\top} \cdot \boldsymbol{\lambda} - \frac{\partial^2 \mathcal{J}}{\partial \boldsymbol{\theta}^2} \cdot \boldsymbol{\xi} \quad , \quad \boldsymbol{\zeta}(0) = \underline{\frac{\partial^2 J}{\partial \boldsymbol{\Theta}^2} \cdot \boldsymbol{r}} = H\boldsymbol{r}$$

output

$$\boldsymbol{\zeta}(T) = 0$$

Procedure of UQ using Second-Order Adjoint Method

1. Estimate an optimum $\boldsymbol{\Theta}^*$ minimizing J based on the adjoint method and a gradient method (we adopt here limited-memory BFGS method)

2. Evaluate the uncertainty of $\boldsymbol{\Theta}^*$ based on the second-order adjoint method and a gradient method (we adopt here the conjugate residual method)

Remarks:

1. An array having size $O(N^2)$ is not needed.
2. Optimum estimation and UQ can be achieved with $O(K)$ computation (K: computation cost needed for a forward computation).

The proposed method is the only one that can estimate both optimum state and its uncertainty even with a system having large degrees of freedom

Second-Order Adjoint of Phase-Field Model

Forward

$$\tau \frac{\partial \hat{\phi}_i}{\partial t} = \epsilon^2 \triangle_i \hat{\phi}_i + \left\{ -3\phi_i^2 + (3-2m)\phi_i + m - \frac{1}{2} \right\} \hat{\phi}_i + \phi_i(1-\phi_i)\hat{m}$$

$$\frac{\partial \hat{m}}{\partial t} = 0$$

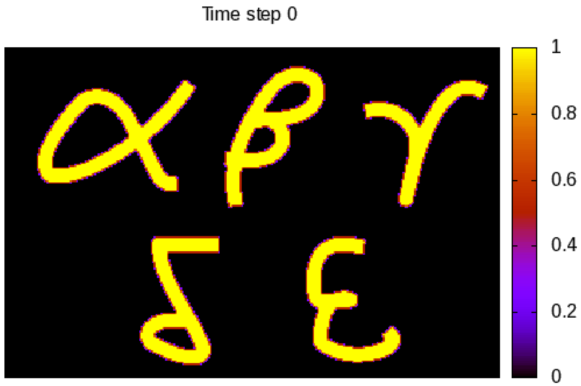
Backward

$$-\tau \frac{\partial \check{\phi}_i}{\partial t} = \epsilon^2 \triangle_i \check{\phi}_i + \left\{ -3\phi_i^2 + (3-2m)\phi_i + m - \frac{1}{2} \right\} \check{\phi}_i + (6\phi_i + 2m - 3)\,\tilde{\phi}_i \check{\phi}_i + (2\phi_i - 1)\,\tilde{\phi}_i \check{m} + \left[\sum_j \frac{\partial^2 \mathcal{J}}{\partial \phi_i \partial \phi_j} \hat{\phi}_j + \frac{\partial^2 \mathcal{J}}{\partial \phi_i \partial m} \hat{m} \right] \sum_{k \in \mathcal{K}} \delta(t - t_k)$$

$$-\tau \frac{\partial \check{m}}{\partial t} = \sum_j \left[\phi_j (1-\phi_j) \check{\phi}_j + (2\phi_j - 1)\,\tilde{\phi}_j \hat{\phi}_j \right] + \left[\sum_j \frac{\partial^2 \mathcal{J}}{\partial m \partial \phi_j} \hat{\phi}_j + \frac{\partial^2 \mathcal{J}}{\partial m^2} \hat{m} \right] \sum_{k \in \mathcal{K}} \delta(t - t_k)$$

Setup of Twin Experiments

Can the proposed method correctly reproduce the true parameter and true initial state from synthetic data that are generated by using the true parameter and initial state?



True initial state

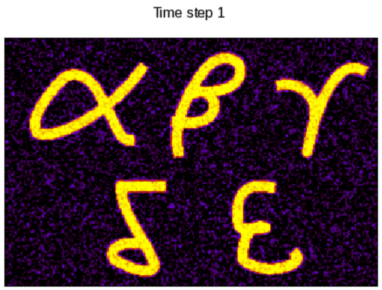
$m_{\mathrm{true}} = 0.1$

Synthetic observation data

True phase field obtained with time interval ΔT

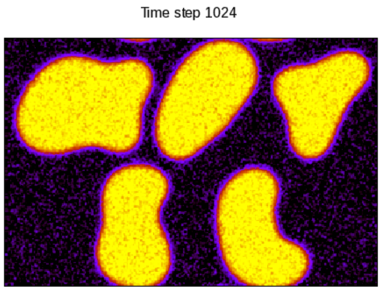
+

Gaussian noise $N\left(0, \sigma^2\right)$



$T_{\mathrm{min}} = 0.1\tau$

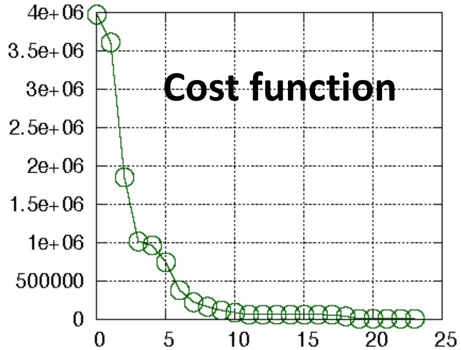
…

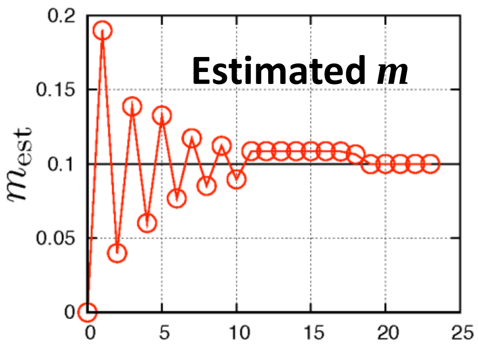


$T_{\mathrm{max}} = 102.4\tau$

$\sigma^2 = 0.01, \quad \Delta T = 0.1\tau$

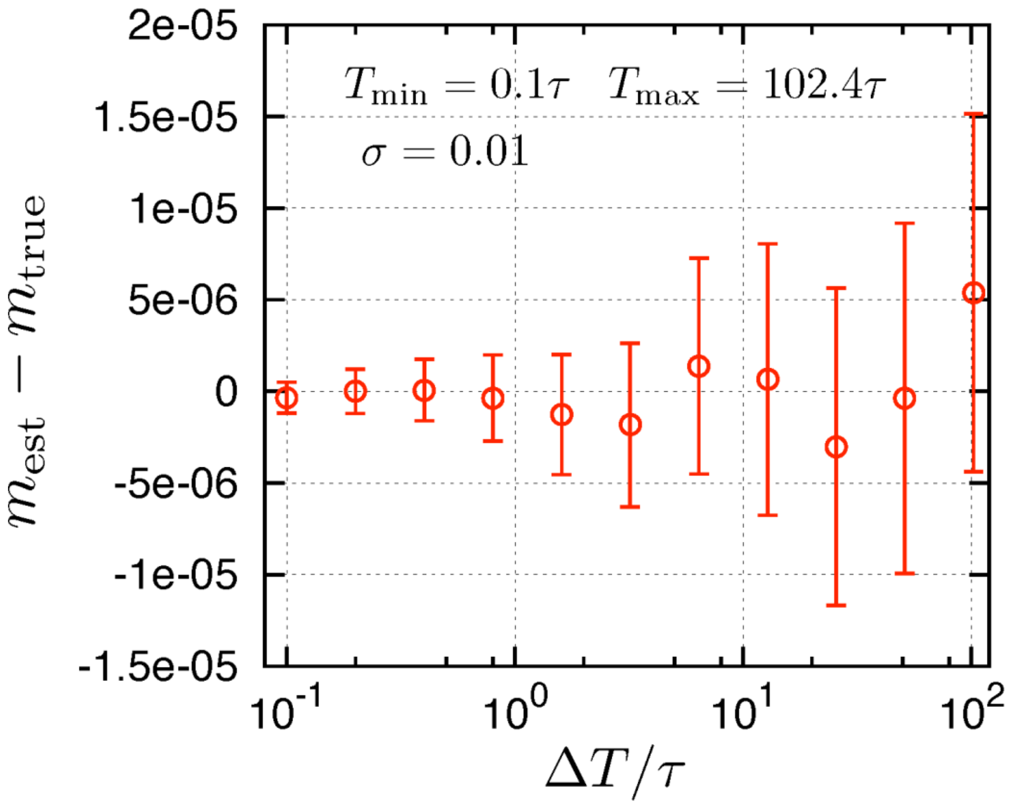
Example of twin experiment

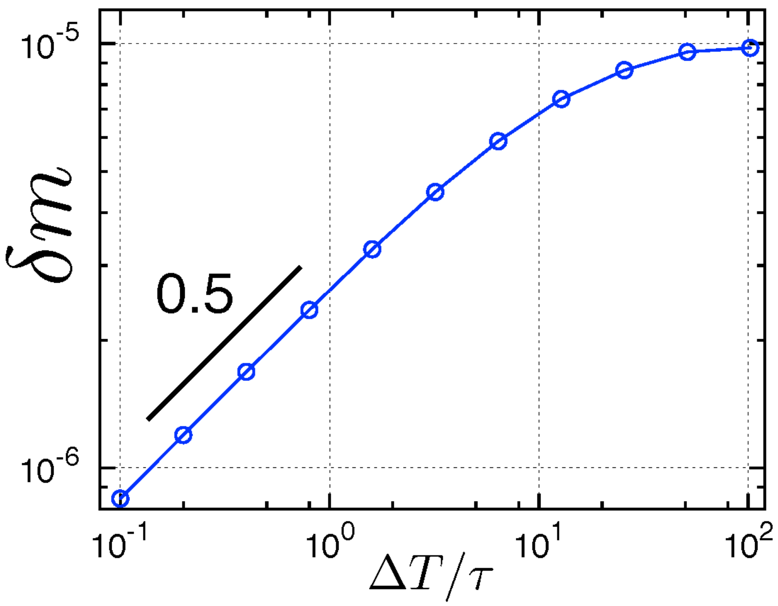




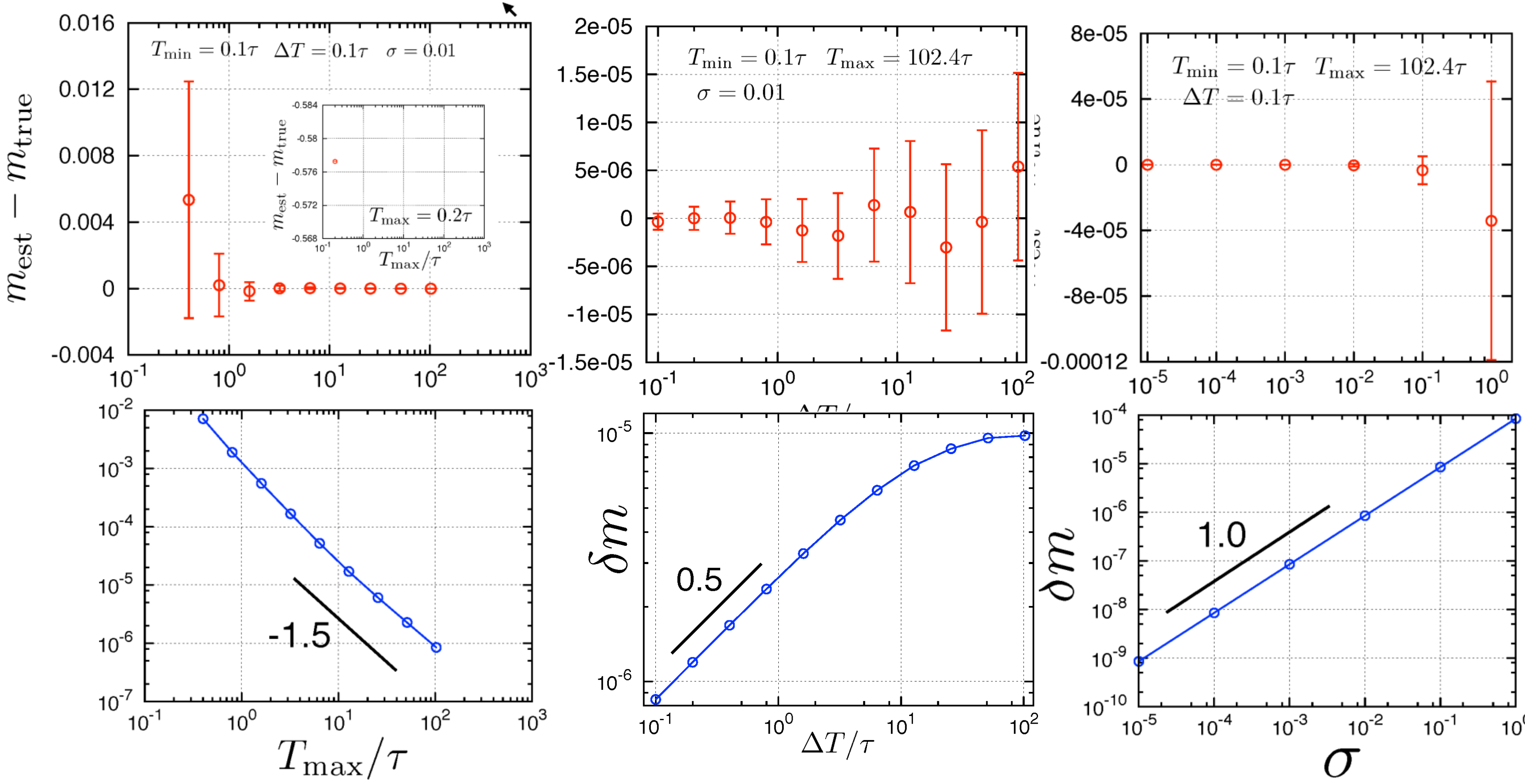
Twin Experiment: Parameter Estimation

How estimate and its uncertainty depend on the time interval of data?





Twin Experiment: Parameter Estimation

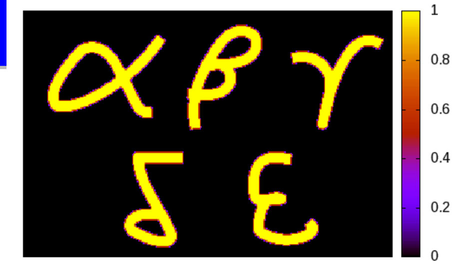


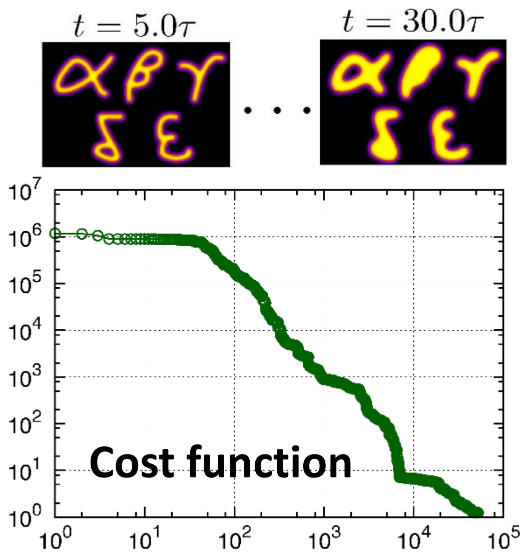
Estimation of parameter and its uncertainty depending on quality and quantity of data

Feedback to experimental design

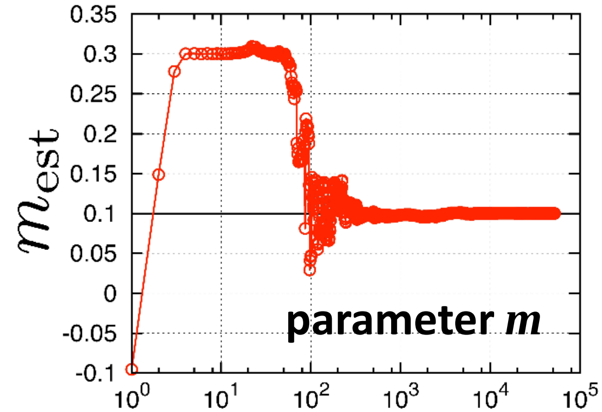
Twin Experiment: Parameter & Initial State

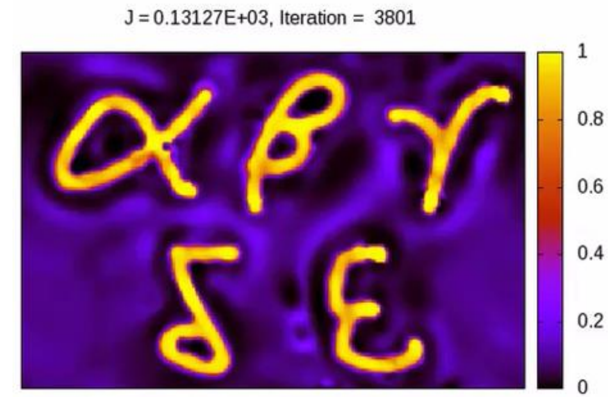
true

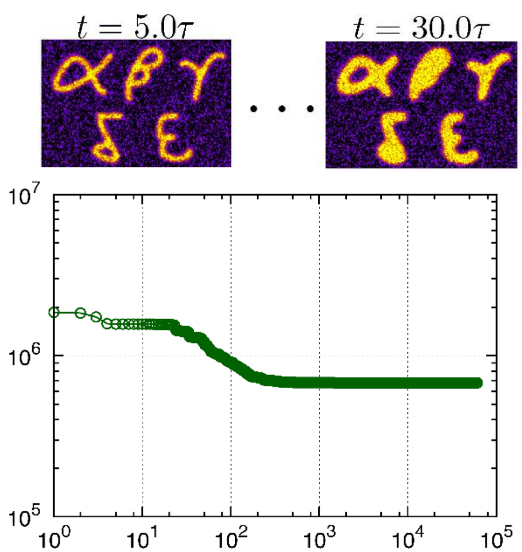




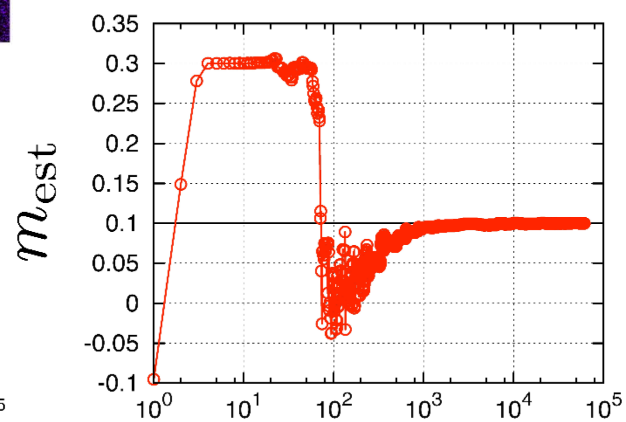
small noise: $\sigma = 1\times10^{-4}$

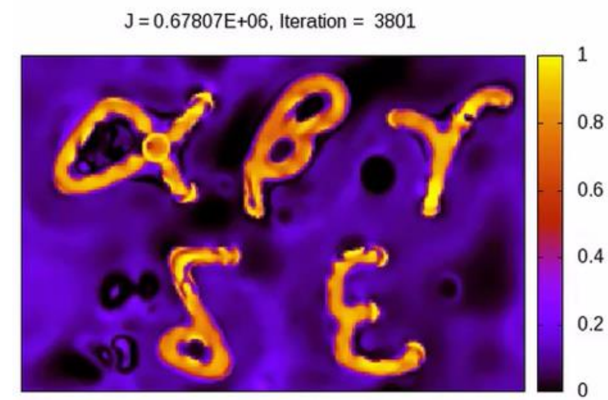






large noise: $\sigma = 0.3$





Summary & Ongoing Works

Summary

1. We have established a DA methodology that enables us to estimate optimum state and parameters but also evaluate their uncertainties based on the second-order adjoint method, which is applicable to a system having large degrees of freedom.
2. Such UQ can give feedback to designs of observations/experiments.

Ongoing (?) works

1. Implement of a Monte-Carlo like method to exclude the dependency of an initial guess
2. Development of a formula manipulation method to derive the first-/second-order derivatives of a given system model, i.e., $\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\theta}}, \frac{\partial^2 \boldsymbol{F}}{\partial \boldsymbol{\theta}^2}$

References

- Ito, S., H. Nagao, A. Yamanaka, Y. Tsukada, T. Koyama, M. Kano, and J. Inoue, Data assimilation for massive autonomous systems based on a second-order adjoint method, Phys. Rev. E, 94, 043307, doi:10.1103/PhysRevE.94.043307, 2016.
- Kano, M., S. Miyazaki, Y. Ishikawa, Y. Hiyoshi, K. Ito, and K. Hirahara, Real data assimilation for optimization of frictional parameters and prediction of afterslip in the 2003 Tokachi-oki earthquake inferred from slip velocity by an adjoint method, Geophys. J. Int., 203, 646-663, doi: 10.1093/gji/ggv289, 2015.
- Kobayashi, R., Modeling and numerical simulations of dendritic crystal growth, Physica D, 63, 410-423, 1993.
- Le Dimet, F.-X., I. M. Navon, and D. N. Daescu, Second-order information in data assimilation, Mon. Weather Rev., 130, 629-648, 2002.
- Maeda, T., K. Obara, M. Shinohara, T. Kanazawa, and K. Uehira, Successive estimation of a tsunami wavefield without earthquake source data: A data assimilation approach toward real-time tsunami forecasting, Geophys. Res. Lett., doi: 10.1002/2015GL065588, 2015.

References

- Shimokawabe, T., T. Aoki, T. Takaki, A. Yamanaka, A. Nukada, T. Endo, N. Maruyama, and S. Matsuoka, Peta scale phase-field simulation for dendritic solidification on the TSUBAME 2.0 Supercomputer, Proceedings of Supercomputing 2011, 1-11, 2011.
- 伊理正夫, 久保田光一, 高速自動微分法(I), 応用数理, 1(1), 17-35, 1991.
- 伊理正夫, 久保田光一, 高速自動微分法(I), 応用数理, 1(2), 53-63, 1991.
- 樋口知之，上野玄太，中野慎也，中村和幸，吉田亮，データ同化入門，朝倉書店，2011.