

サンプリングの高速化へ向けた4次元変分法の発展と応用

伊藤 伸一

東京大学 地震研究所
東京大学大学院 情報理工学系研究科

スマートサンプリング講演会 2021.01.22

自己紹介

学歴・職歴

2015.03 大阪大学 大学院理学研究科 博士（理学）

- ・破壊のパターン形成の数理

2015.04–2018.08 東京大学 地震研究所 特任研究員

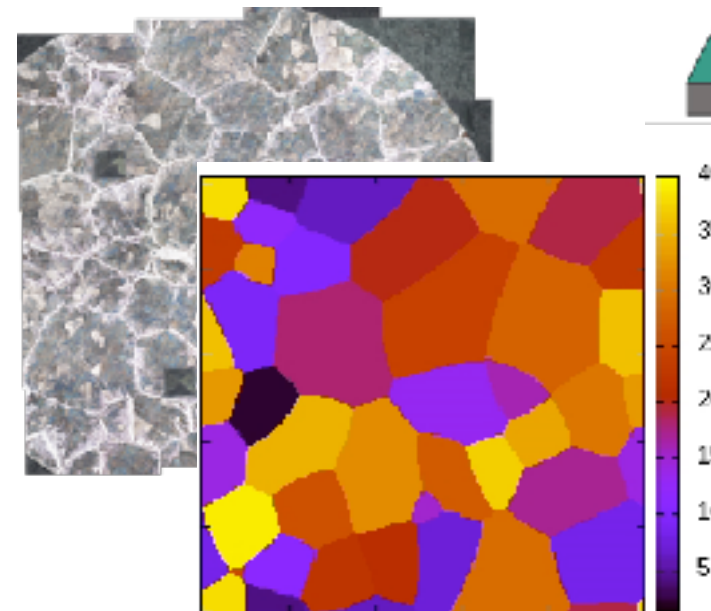
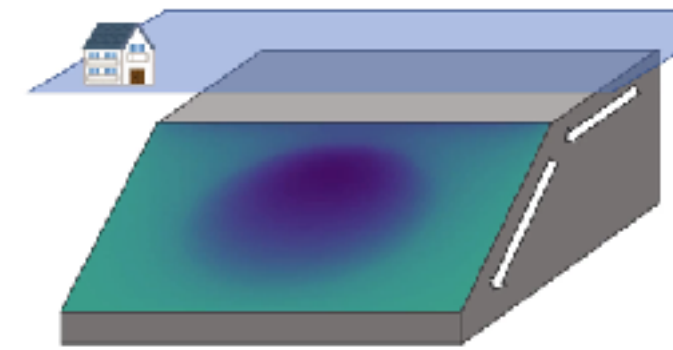
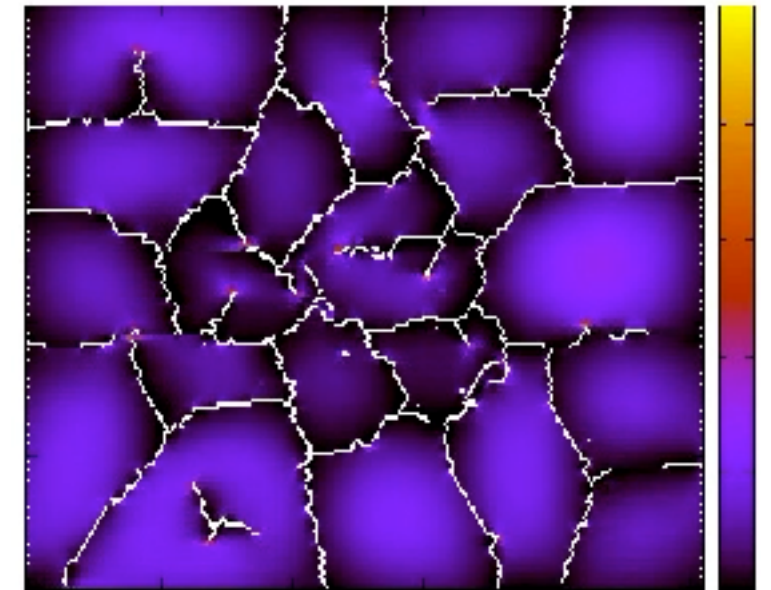
2018.08–現在 東京大学 地震研究所 助教

2018.12–現在 東京大学 大学院情報理工学系研究科 助教（兼務）

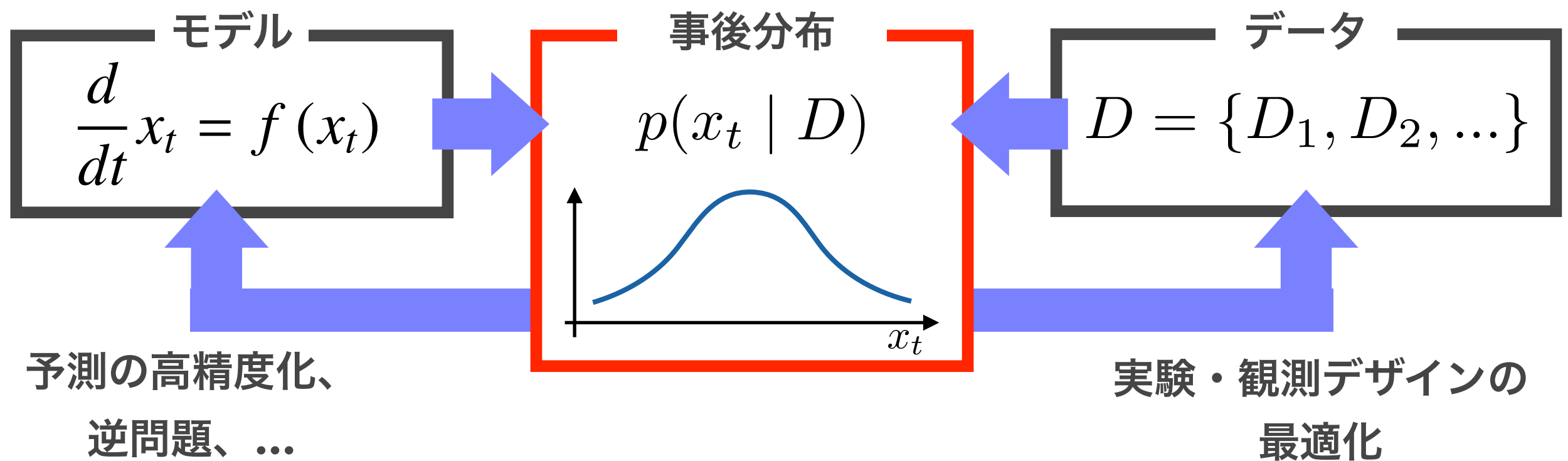
- ・大規模4次元変分法データ同化の数理・応用

- ・金属材料の粒成長パターン

- ・断層摩擦パラメータの空間パターン



データ同化



事後分布（に比例した量）を1点評価するのに1回のモデルの時間発展計算が必要

- ・ 小さいモデルを使う？ → 表現能力の低下
- ・ 評価回数を極力減らす？ → 事後分布の精度低下

→ モデルの規模・目的に適したアルゴリズムの選択が大切

本日の話題

アルゴリズムの話題を中心にお話しします。

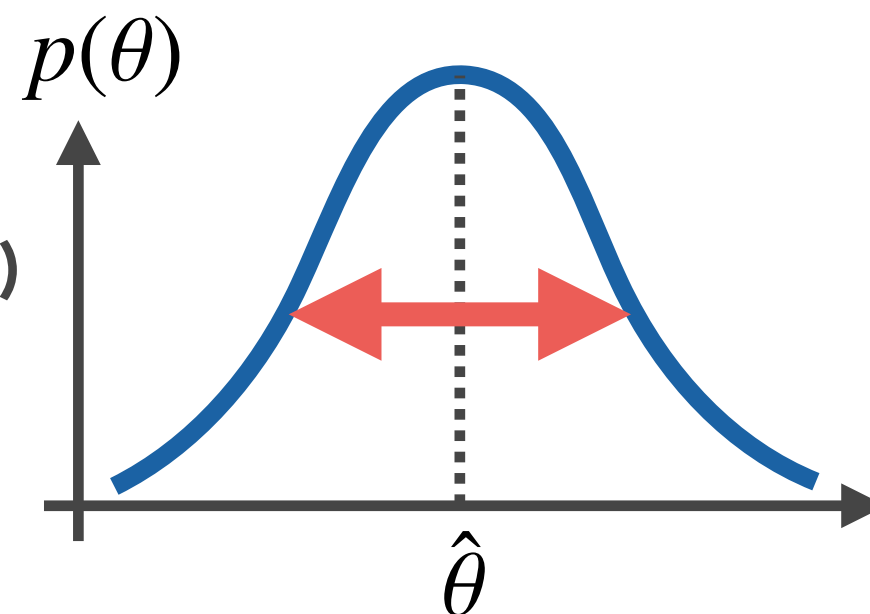
- 4次元変分法データ同化
- サンプルング法との融合可能性
相補的な高速化へ向けた試み（模索中）

休憩

- 大規模不確実性評価

Ito et al., Phys. Rev. E, 94, 043307 (2016)

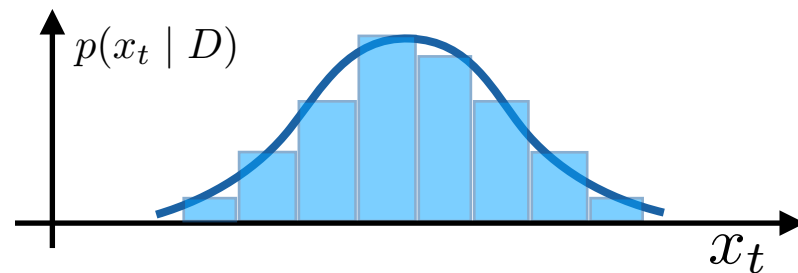
Ito et al., STAM, 18:1, 857-869 (2017)



4次元変分法データ同化と サンプリング法との融合可能性

データ同化アルゴリズムあれこれ

逐次データ同化



- 概形と時間発展を評価

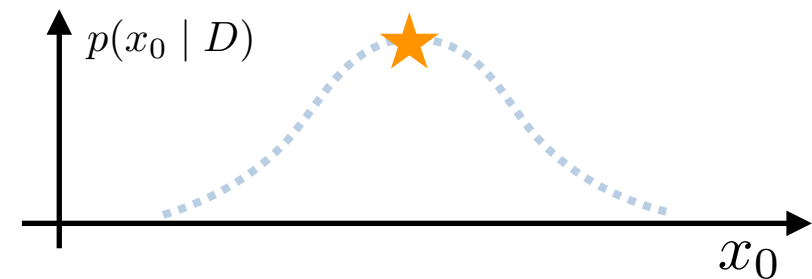
事後分布をサンプル群の時間発展で評価

- 小中規模モデル向き

たくさんのシミュレーションを同時に走らせる必要がある（並列化可）

- EnKF, 粒子フィルタ など

非逐次データ同化



- 概形も時間発展も評価しない

初期値の事後分布のMAP解だけを探索

- 大規模モデル向き

シミュレーション1つに
計算資源を集中できるため

- 4次元変分法

大規模モデル

偏微分方程式

$$\frac{\partial y}{\partial t} = \nabla (c \nabla y)$$
$$\frac{\partial c}{\partial t} = 0$$

空間方向
離散化

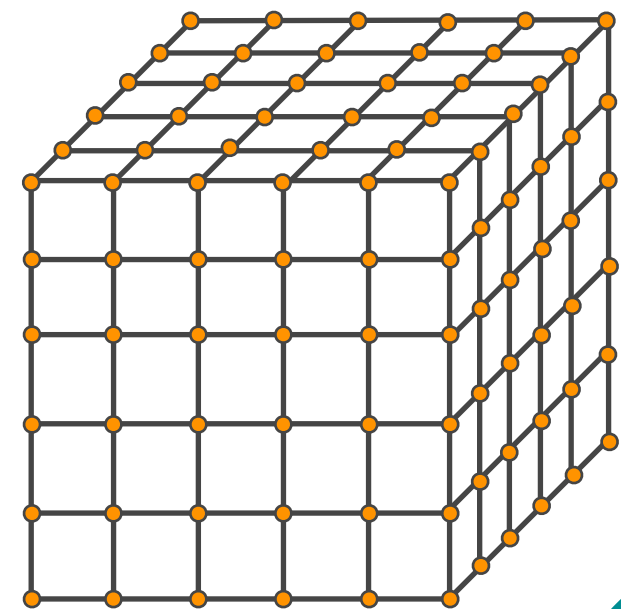
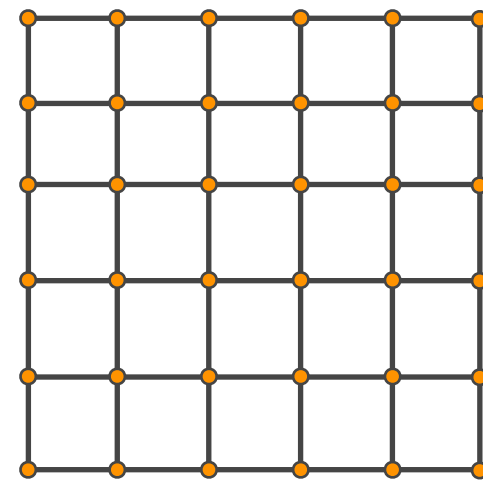


常微分方程式

$$\frac{dx}{dt} = F(x)$$

$$x = (y_1, y_2, \dots, c_1, c_2, \dots)^T$$

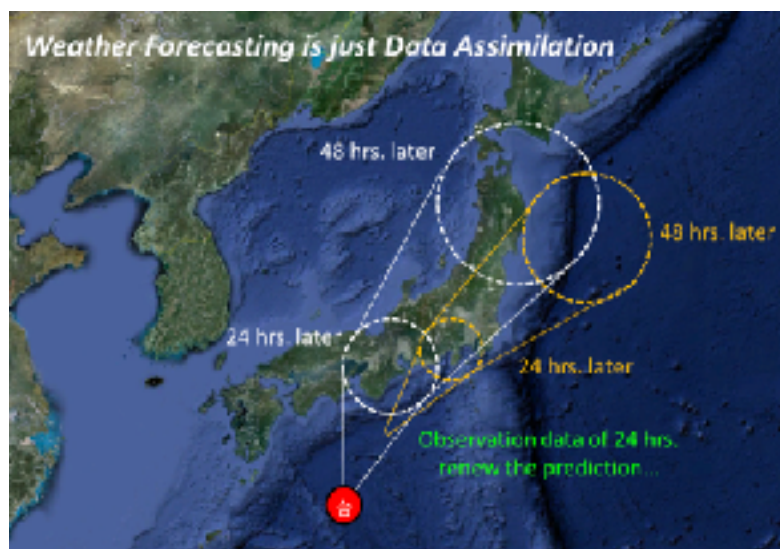
↑ 格子点上の物理量を全て並べたベクトル



空間1方向に100格子点とると,

2次元問題 100x100 格子点 = $O(10,000)$ 変数!

3次元問題 100x100x100 格子点 = $O(1,000,000)$ 変数!!

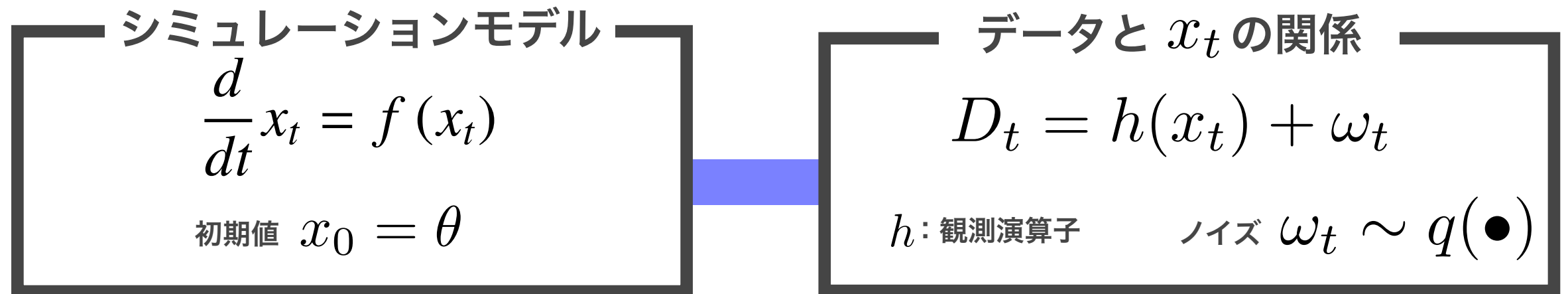


出典: Google Earth

大規模モデルの事後分布を
効率よく評価するアルゴリズム

4次元変分法

モデルの初期値に関する事後分布から事後確率最大解（MAP解）を抽出する手法

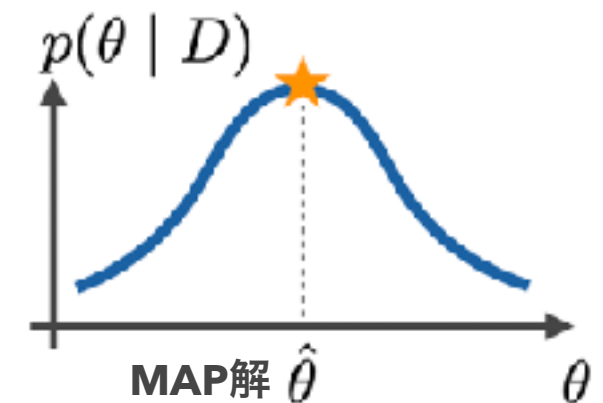


事後分布

$$p(\theta | D) \propto p(\theta) \prod_{t \in T^{\text{obs}}} q(D_t - h(x_t))$$

$p(\theta)$: 事前分布

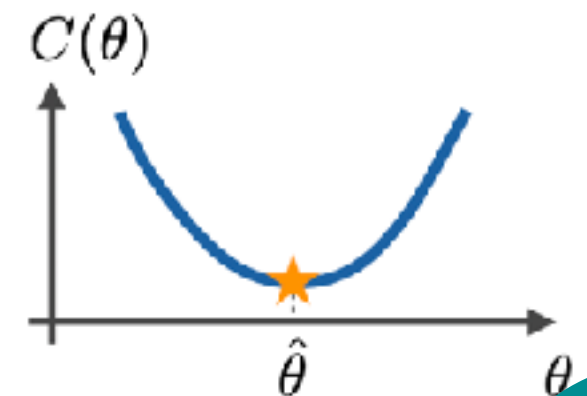
T^{obs} : 観測時刻のセット



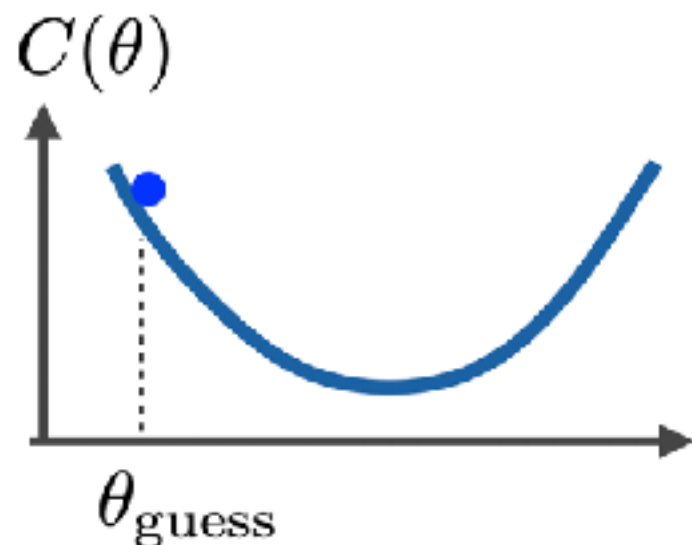
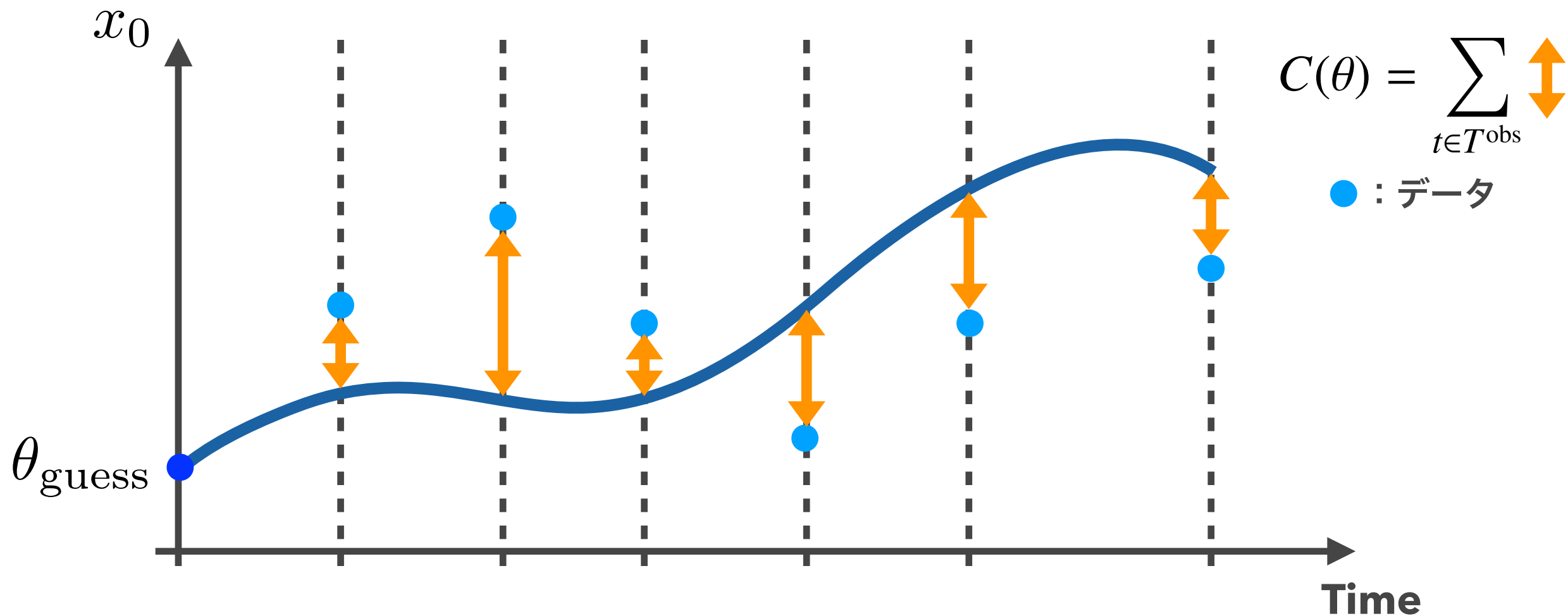
コスト関数

$$C(\theta) = -\log p(\theta) - \sum_{t \in T^{\text{obs}}} \log q(D_t - h(x_t))$$

$$p(\theta | D) \propto e^{-C(\theta)}$$

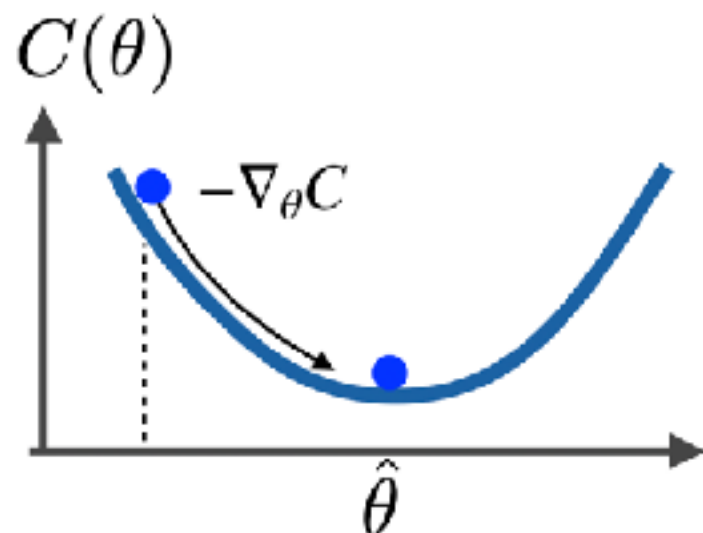
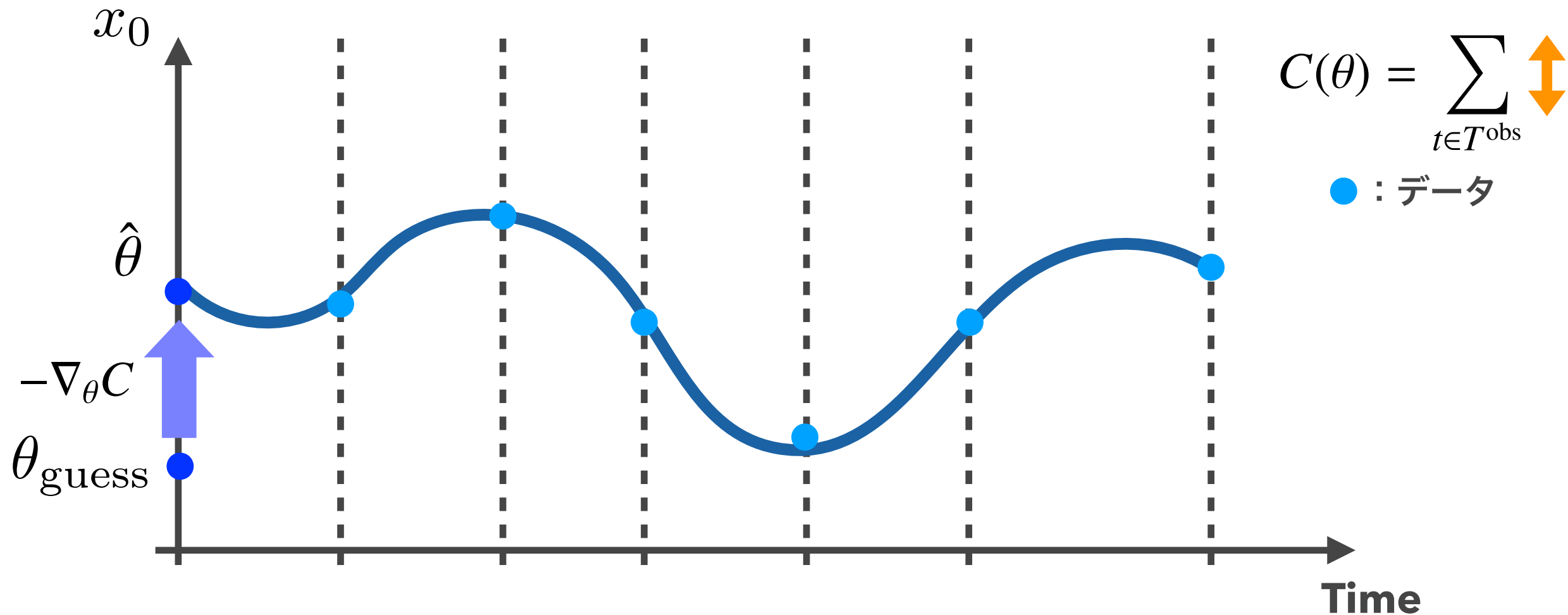


4次元変分法による初期値更新



適当に選んだ $x_0 = \theta_{\text{guess}}$ ではデータと全く合わないが...

4次元変分法による初期値更新



勾配法による最適化により
データに適合する初期値 $x_0 = \hat{\theta}$ を探す

➡ $C(\theta)$ の勾配 $\nabla_{\theta} C$ の計算が必要

Adjoint法による勾配計算

勾配法によって最適化したいが...

$$C(\theta) = -\log p(\theta) - \sum_{t \in T^{\text{obs}}} \log q(D_t - h(x_t))$$

θ で直接微分ができない

Adjoint法

$$L = C + \int_0^T dt \lambda_t^\top \left(f(x_t) - \frac{d}{dt} x_t \right)$$

T : 終端時刻
 λ_t : **Adjoint**変数
(ラグランジュの未定乗数)

変分をとって係数比較

Adjoint モデル

$$-\frac{d}{dt} \lambda_t = (\nabla_x f)^\top \lambda_t + \sum_{t' \in T^{\text{obs}}} \delta(t - t') \nabla_{x_{t'}} C$$

$\lambda_0 = \nabla_\theta C$
 $\lambda_T = 0$

t=Tから時間後る向きに解くことで、目的の勾配が得られる

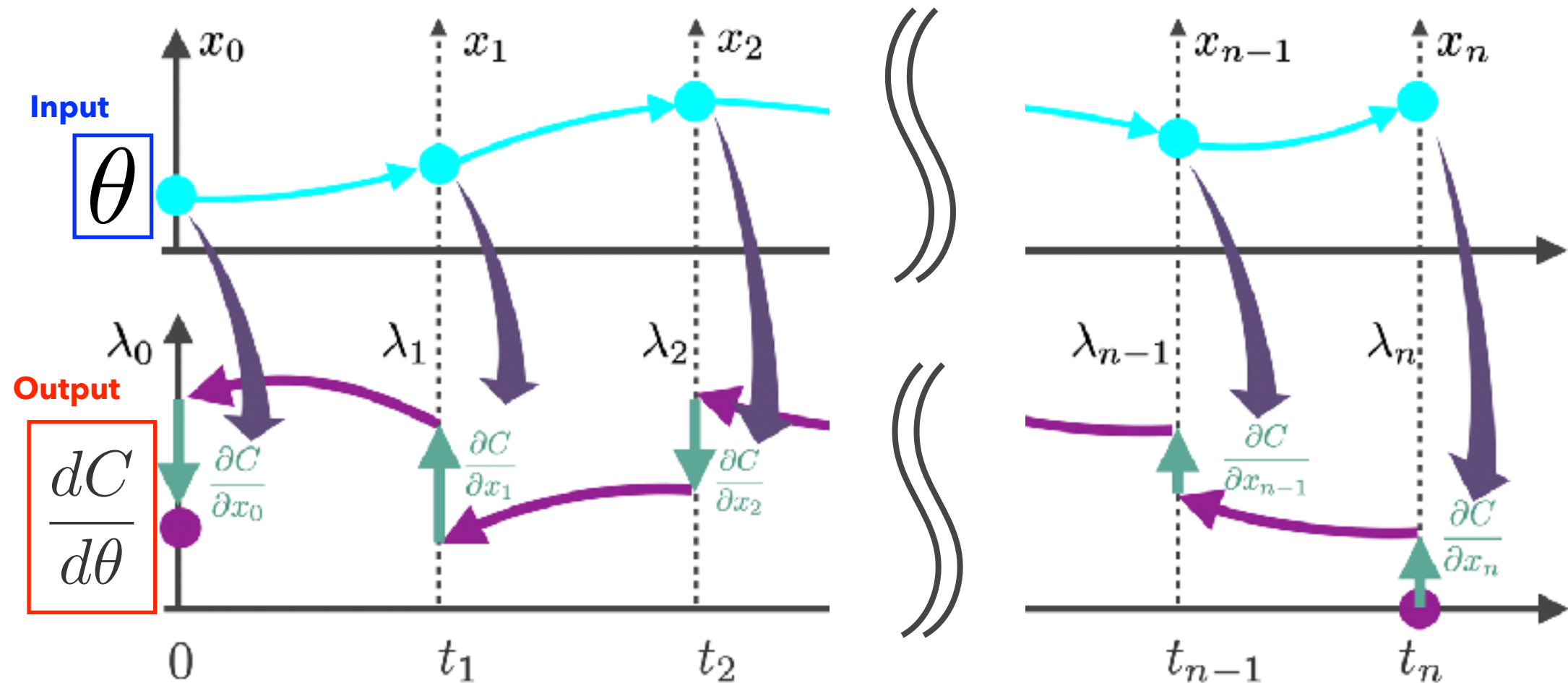
時間後ろ向きに計算する

Adjoint モデル

$$-\frac{d}{dt}\lambda_t = (\nabla_x f)^\top \lambda_t + \sum_{t' \in T^{\text{obs}}} \delta(t - t') \nabla_{x_{t'}} C$$

$$\lambda_0 = \nabla_{\theta} C$$

$$\lambda_T = 0$$



$$T^{\text{obs}} = \{0, t_1, t_2, \dots, t_n = T\}$$

4次元変分法

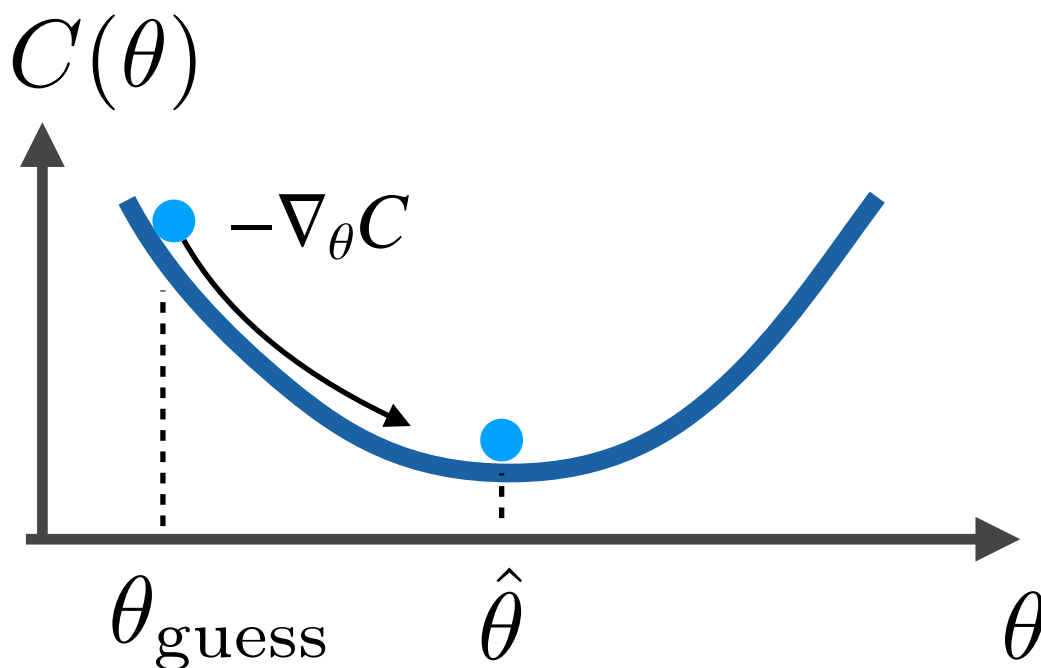
Adjoint モデル

$$-\frac{d}{dt}\lambda_t = (\nabla_x f)^\top \lambda_t + \sum_{t' \in T^{\text{obs}}} \delta(t - t') \nabla_{x_{t'}} C$$

$$\lambda_0 = \nabla_\theta C$$

$$\lambda_T = 0$$

4次元変分法による初期値最適化



- ① 初期推定値を設定
- ② シミュレーションモデルを解く
- ③ Adjoint モデルを時間後る向きに解く
- ④ 勾配に従って少し動かす。①へ戻る。

Adjointモデルの計算コスト ≡ シミュレーションモデルの計算コスト

➡ 全体の計算コスト = (シミュレーション時間) × (反復回数)

c.f. 数値微分 (シミュレーション時間) × (変数の数) × (反復回数)

4次元変分法の適用：反応拡散系

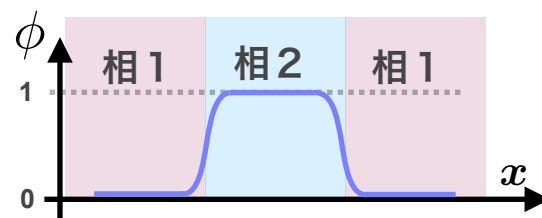
Allen-Cahn モデル (phase-field モデル)

2つの相の時空間発展を記述

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

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

$\phi(x, t)$: Phase-field 変数

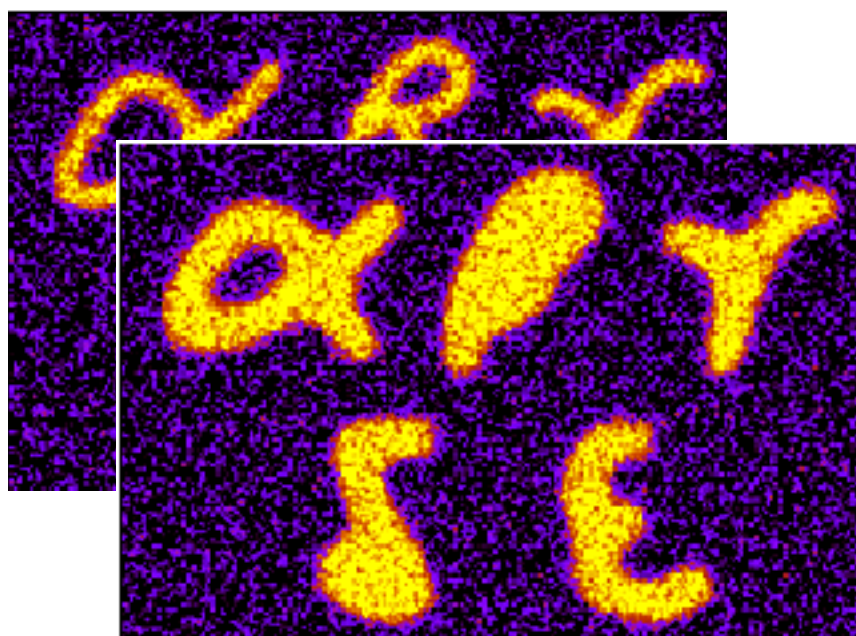


m : 相境界移動方向を決めるパラメータ

TDGL方程式、結晶成長モデル、延焼モデル、...



問題設定



Noisy なスナップショットの時系列データから

- ϕ の初期値 $\phi(x, 0)$
- m の値

を4次元変分法によって推定 (**60,001次元**)

詳細な設定

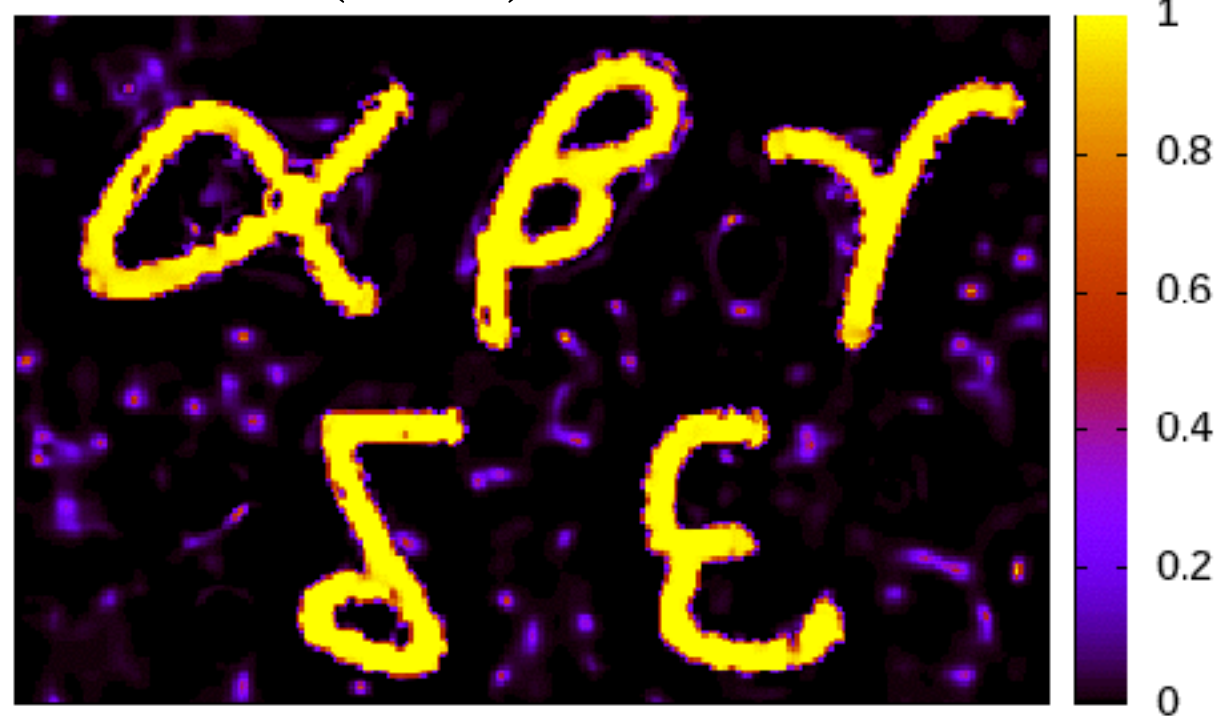
ノイズ分布: $q(\omega) = N(0, 0.3^2)$ 事前分布: 一様分布

データ取得: 時間窓 $[5\tau, 30\tau]$ を 0.1τ 間隔

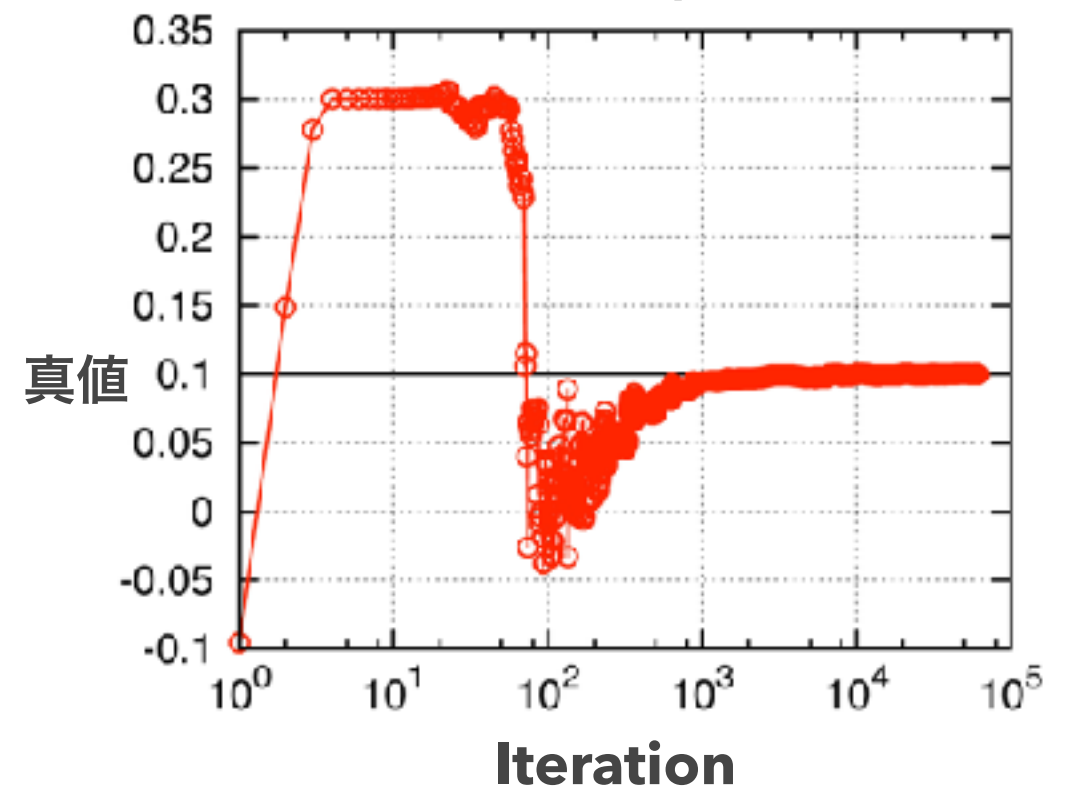
最適化法: L-BFGS-B 法

4次元変分法の適用：反応拡散系

$\phi(x, 0)$ の推定



m の推定



4次元変分法のまとめ

良いところ

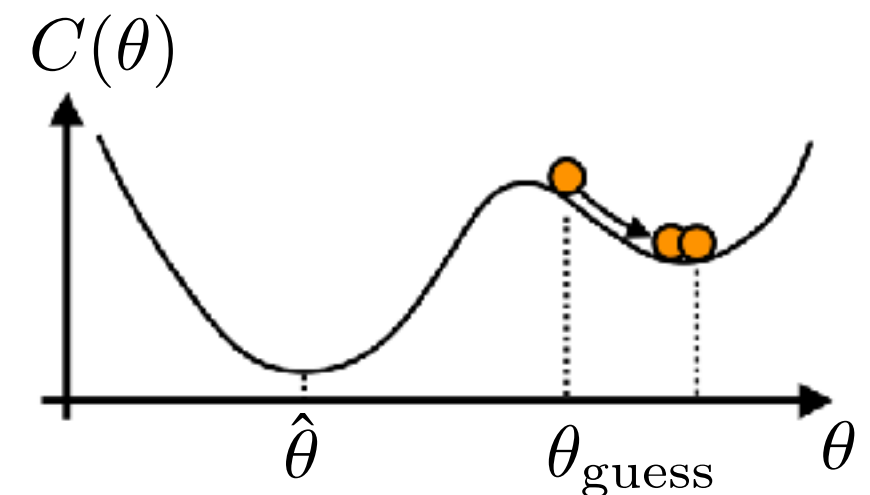
- ・モデルの初期値に関する事後分布を評価
- ・事後確率最大解（MAP解）のみを探索

→ 大規模モデルのパラメータ推定・初期値問題に非常に強い

改善したいところ

- ・多峰性による局所解への落ち込み
 - ・通常の4次元変分法では抜け出せない
- ・より高速な探索
 - ・モデル計算回数はなるべく少なくしたい

➡ サンプルングの手法と組み合わせてみる



Langevin dynamics

- 通常の4次元変分法

$$\theta_{k+1} = \theta_k - \tau \nabla_{\theta_k} C$$

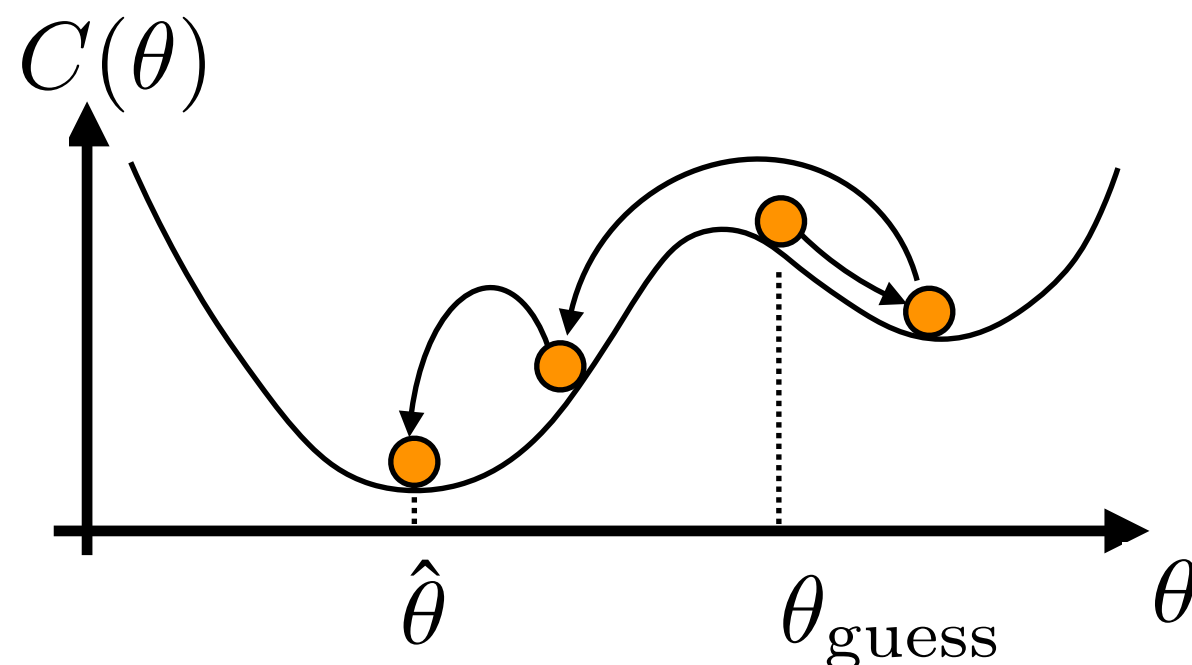
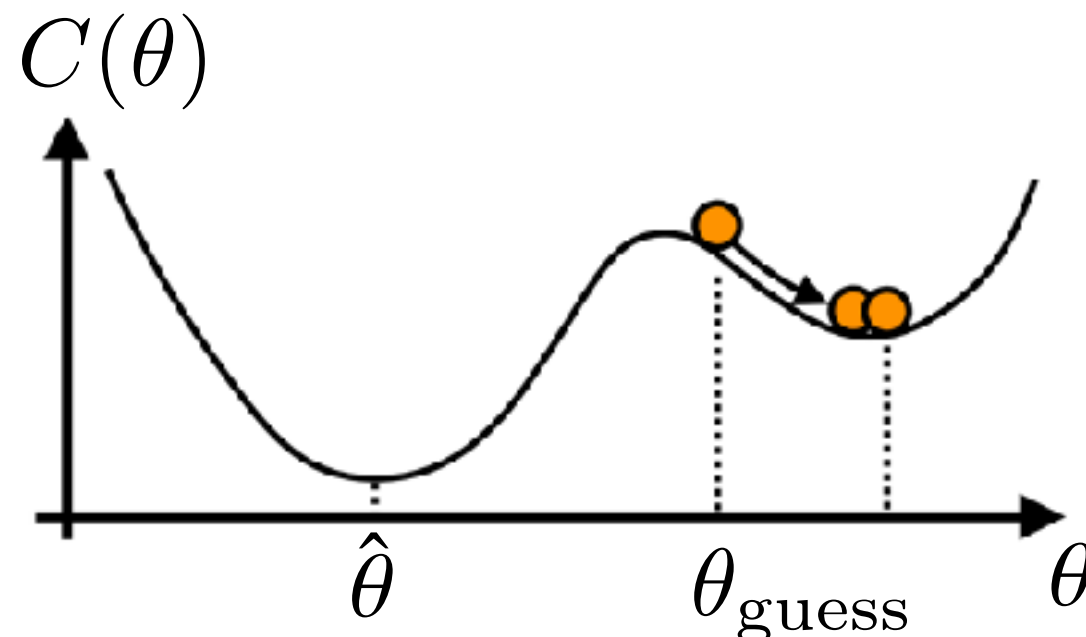


- Metropolis-adjusted Langevin algorithm

$$\theta_{k+1} = \theta_k - \tau \nabla_{\theta_k} C + \sqrt{2\tau} \xi_k$$

ξ_k : ガウスノイズ

提案分布 $q(\theta' | \theta) \propto \exp \left[\frac{1}{4\tau} \|\theta' - \theta + \tau \nabla_{\theta} C\|_2^2 \right]$



- 4次元変分法 → 局所解からの脱出
- サンプリング法 → 勾配によるブースト

例題

- 特徴的な粒サイズ

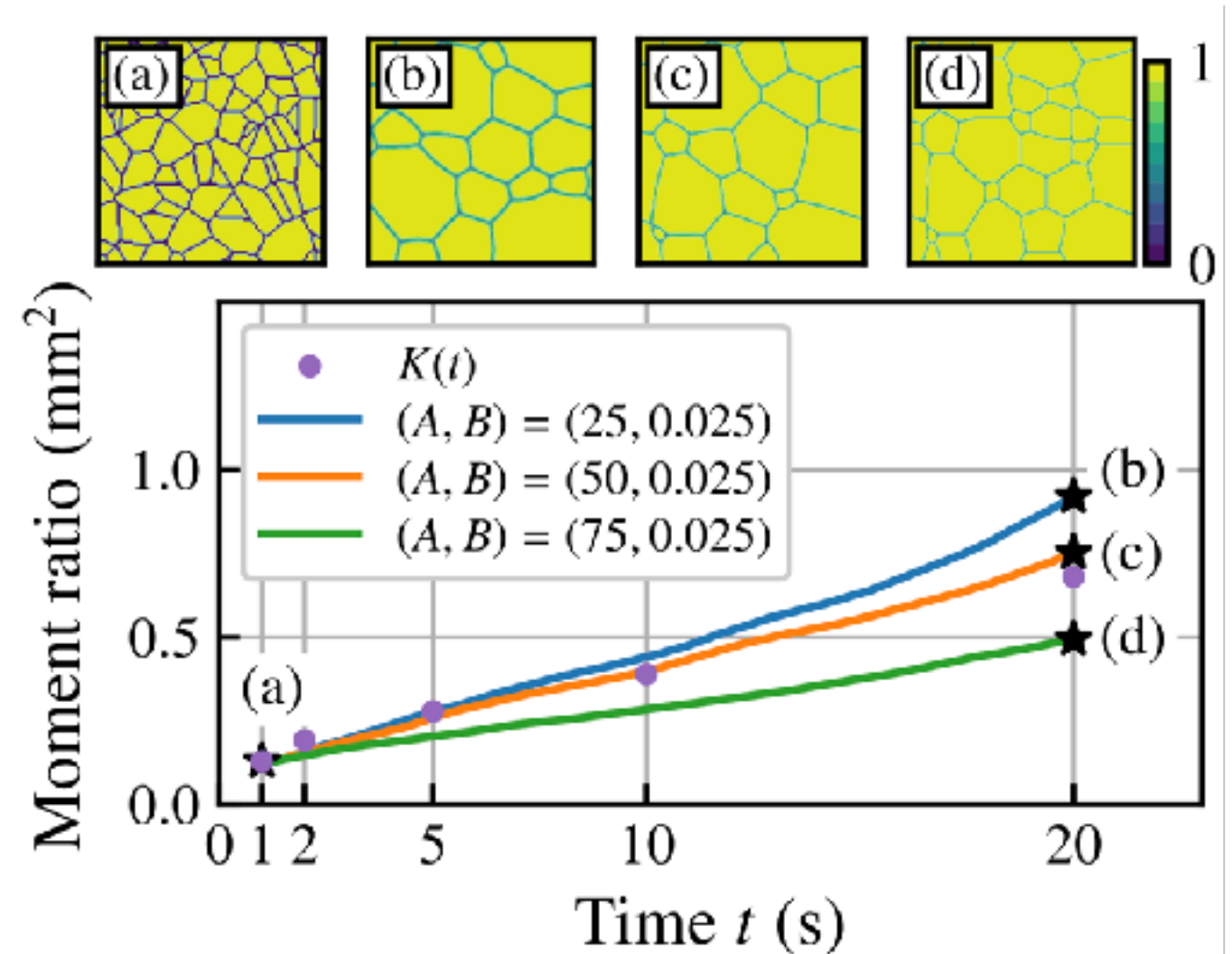
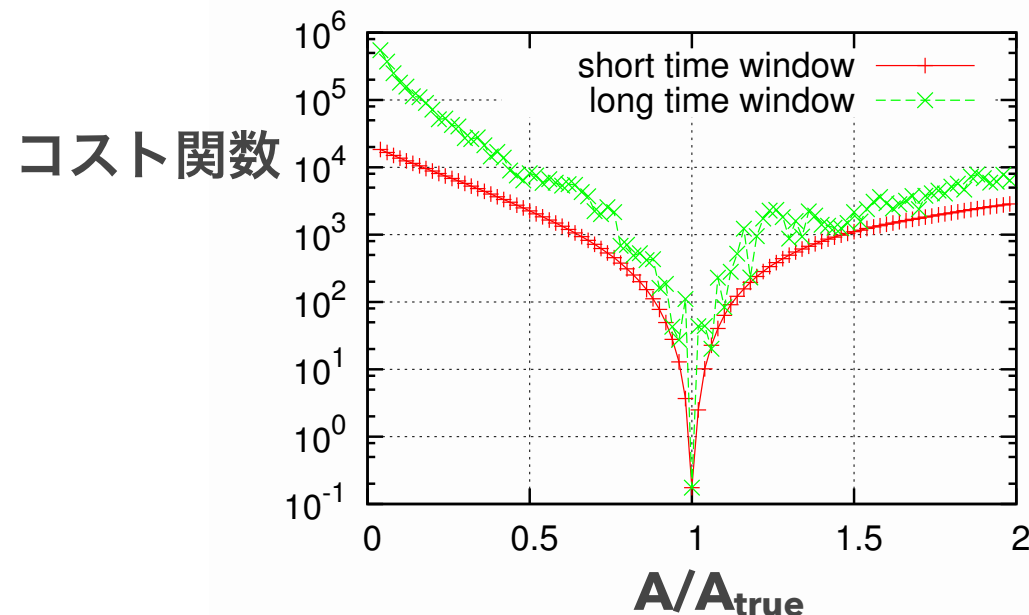
$$\mathcal{K}(t) = \frac{\sum_{i=1}^N S_i^2(t)}{\sum_{i=1}^N S_i(t)}$$

where $S_i(t) = \int d\mathbf{x} \Psi_i(\mathbf{x}, t)$

- 擬似観測データ

$$K(t) = \mathcal{K}(t) + \omega_t$$

ω_t : ガウスノイズ



- コスト関数がガタガタになることがある
→ 通常の4次元変分法が使えない

$$\theta_{k+1} = \theta_k - \tau \nabla_{\theta_k} C + \sqrt{2\tau} \xi_k$$

ノイズの力を借りて局所解から脱出

レプリカ交換モンテカルロ法

逆温度 β

高温

山越えが容易



山越えが困難

低温

温度の異なる分布族

$$p_{\beta_i}(\theta \mid D) \propto e^{-\beta_i C(\theta)}$$

$$0 = \beta_1 < \beta_2 < \dots < \beta_n = 1$$

$\beta_n = 1$ が元々の事後分布

1. それぞれは独自にマルコフ連鎖で更新
2. 時々異なる温度間でサンプルを交換

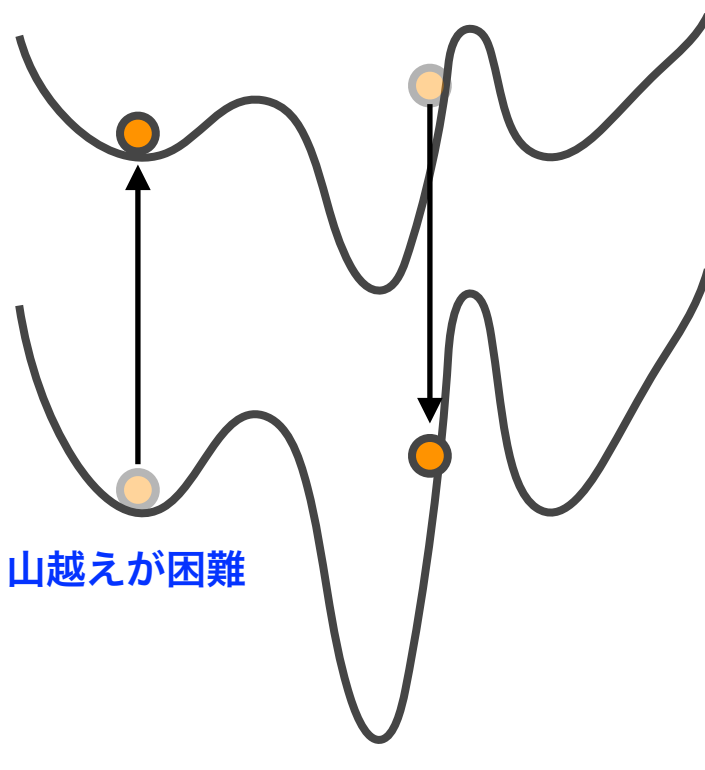
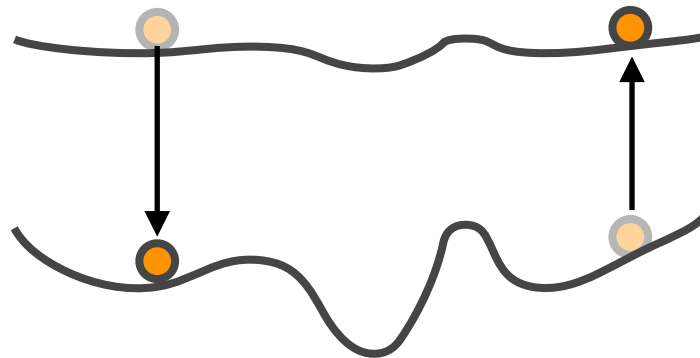
単純なメトロポリス法では超えられない
大きな障壁を越えることが可能

レプリカ交換モンテカルロ法

逆温度 β

高温

山越えが容易



山越えが困難

低温

温度の異なる分布族

$$p_{\beta_i}(\theta \mid D) \propto e^{-\beta_i C(\theta)}$$

$$0 = \beta_1 < \beta_2 < \dots < \beta_n = 1$$

$\beta_n = 1$ が元々の事後分布

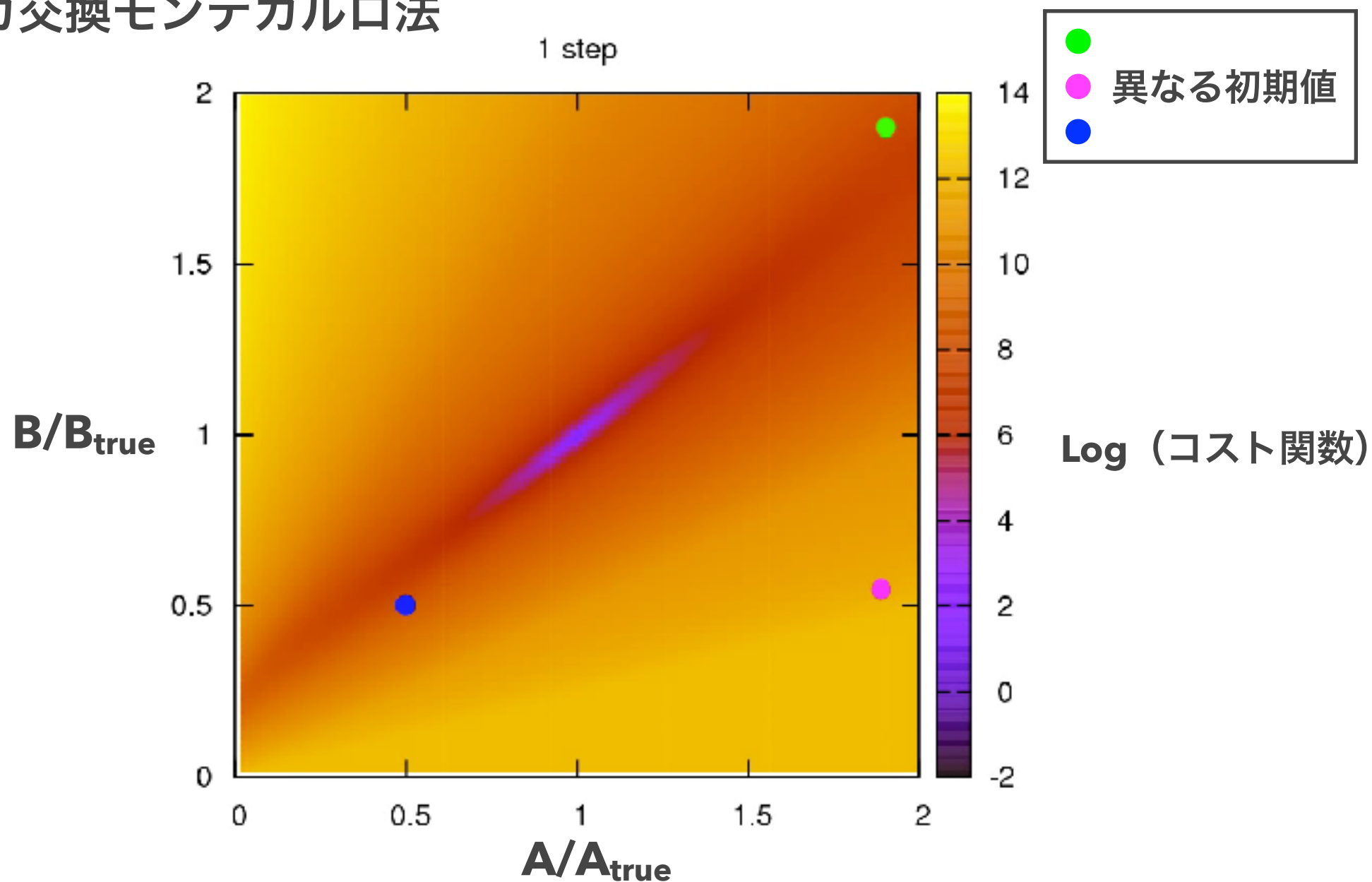
1. それぞれは独自にマルコフ連鎖で更新
2. 時々異なる温度間でサンプルを交換

単純なメトロポリス法では超えられない
大きな障壁を越えることが可能

数値実験

Metropolis-adjusted Langevin algorithm

+ レプリカ交換モンテカルロ法



初期値に依らず最適解近くに高速に到達

前半のまとめ

4次元変分法

- モデルの初期値に関する事後分布のMAP解を高速に評価 ※良い initial guess を選択できれば
大規模モデルのパラメータ推定・初期値問題に非常に強い

- サンプリング法との融合可能性

- Metropolis-adjusted Langevin algorithm

- Hamiltonian/Hybrid Monte Carlo (HMC)

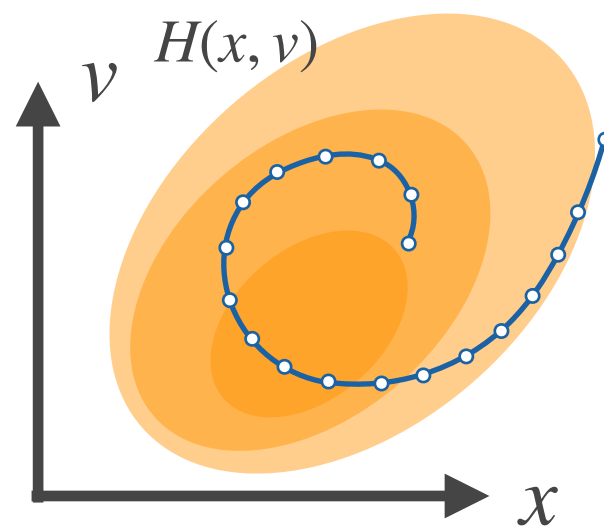
← レプリカ交換法との併用

- Hamiltonian/Hybrid Monte Carlo (HMC)

$$H(x, v) = \frac{v^2}{2} - \log(P(x))$$

$$\frac{dv}{dt} = -\frac{\partial H}{\partial x} \quad \frac{dx}{dt} = \frac{\partial H}{\partial v}$$

$e^{-H(v,x)}$ からのサンプリング
→ 採択率が向上



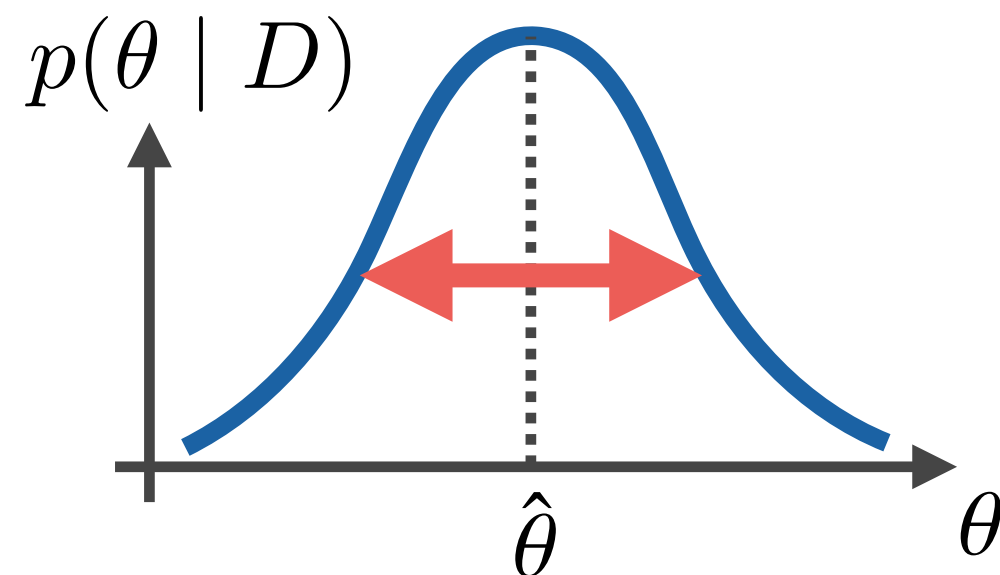
最適化法とサンプリング法
両者の相補的な高度化

後半の話題

もっと改善したいところ

- 4次元変分法単体では事後分布の形状評価は**原理的に不可能**
 - 分散くらいは評価できないか？
 - 推定の不確実性・信頼性評価
 - サンプルング法との組み合わせ？
 - 計算回数をなるべく減らしたい

➡ ラプラス近似による不確実性評価
Ito et al. (2016)



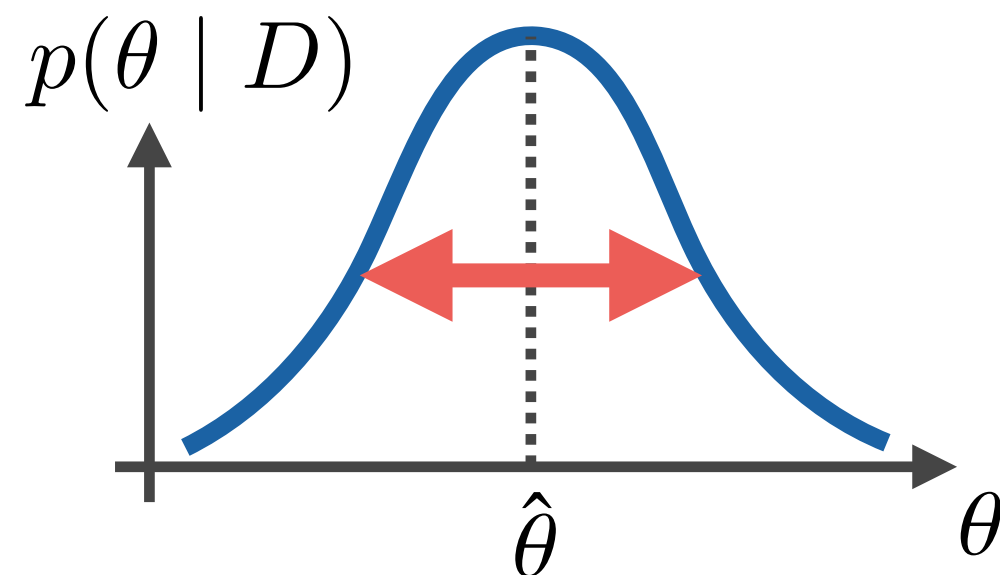
大規模不确实性評価

後半の話題

もっと改善したいところ

- 4次元変分法単体では事後分布の形状評価は**原理的に不可能**
 - 分散くらいは評価できないか？
 - 推定の不確実性・信頼性評価
 - サンプルング法との組み合わせ？
 - 計算回数をなるべく減らしたい

➡ ラプラス近似による不確実性評価
Ito et al. (2016)

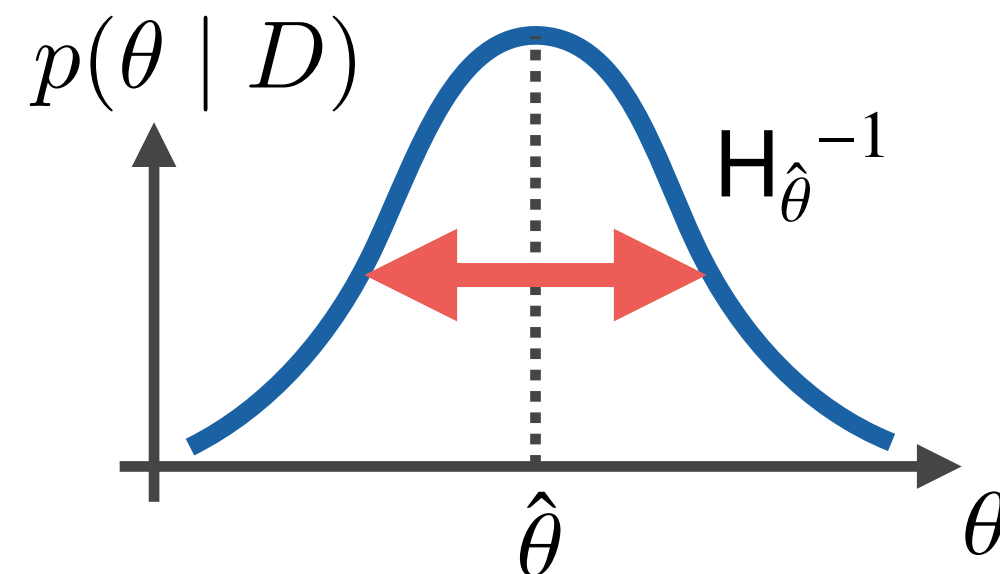


ラプラス近似

$$C(\theta) \sim C(\hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})^\top H_{\hat{\theta}}(\theta - \hat{\theta}) \quad H_{\hat{\theta}} = \nabla_{\hat{\theta}} \nabla_{\hat{\theta}}^\top C : \text{ヘッセ行列}$$

↓

$$p(\theta | D) \sim \sqrt{\frac{\det H_{\hat{\theta}}}{(2\pi)^N}} \exp \left[-\frac{1}{2}(\theta - \hat{\theta})^\top H_{\hat{\theta}}(\theta - \hat{\theta}) \right]$$



→ ヘッセ行列/逆行列の大きさは (変数の数)²

→ 直接法に基づいてヘッセ逆行列を計算しようとする

$$\text{計算コスト} = \underbrace{(\text{シミュレーション時間}) \times (\text{変数の数})^2}_{\text{要素計算}} + \underbrace{(\text{変数の数})^3}_{\text{逆行列計算}}$$

ヘッセ行列/逆行列の直接計算は避けた

クリロフ部分空間法

ヘッセ逆行列要素を部分的に抽出することを考える

$$Hy = b \quad b = e_i : \text{単位行列の列ベクトル に取れば}$$
$$y_i = (H_\theta^{-1})_{i,i} \text{ は分散、 } y_j = (H_\theta^{-1})_{i,j} \text{ は共分散}$$

クリロフ部分空間法

$$\mathcal{K}_r(H, b) = \text{span}\{b, Hb, H^2b, \dots, H^{r-1}b\}$$

- 共役勾配(CG)法
- 共役残差(CR)法
- 一般化最小残差(GMRES)法,...

CR法の疑似コード

Set an initial guess r_0

Compute $s_0 = b - Hr_0, p_0 = r_0$

For $k = 0, 1, 2, \dots$

$$\alpha_k = (\underline{Hs_k}, s_k) / (\underline{Hs_k}, \underline{Hs_k})$$

$$r_{k+1} = r_k + \alpha_k p_k$$

$$s_{k+1} = s_k - \alpha_k \underline{Hp_k}$$

$$\beta_k = (\underline{Hs_{k+1}}, s_{k+1}) / (\underline{Hs_k}, s_k)$$

$$p_{k+1} = s_{k+1} + \beta_k p_k$$

$$\underline{Hp_{k+1}} = \underline{Hs_{k+1}} + \beta_k \underline{Hp_k}$$

End For

- ヘッセ行列-ベクトル積が高速に計算できれば
逆行列要素が高速に抽出可能になる

Second-order adjoint (SOA) 法

シミュレーションモデルおよび Adjoint モデルを各変数について線形化する

シミュレーションモデル

$$\frac{d}{dt}x_t = f(x_t) \quad x_0 = \theta$$

Adjoint モデル

$$-\frac{d}{dt}\lambda_t = (\nabla_x f)^\top \lambda_t + \sum_{t' \in T^{\text{obs}}} \delta(t - t') \nabla_{x_{t'}} C$$

$$\lambda_T = 0 \quad \lambda_0 = \nabla_\theta C$$

Second-order adjoint (SOA) 法

シミュレーションモデルおよび Adjoint モデルを各変数について線形化する

Tangent linear (TL) モデル

$$\frac{d}{dt}\delta_t = (\nabla_x f) \delta_t \quad \delta_0 = \underline{r}$$

Second-order adjoint (SOA) モデル

$$-\frac{d}{dt}\xi_t = (\nabla_x f)^\top \xi_t + \{(\nabla_x \nabla_x f) \delta_t\}^\top \lambda_t + \sum_{t' \in T^{\text{obs}}} \delta(t - t') \nabla_{x_t} \nabla_{x_t}^\top C \delta_t$$

$$\xi_T = 0 \quad \xi_0 = H_\theta \delta_0 = \underline{H_\theta r}$$

TLモデルとSOAモデルを組み合わせることで、
ヘッセ行列と任意のベクトルの積をベクトルとして出力することができる！

SOA 法によるヘッセ行列-ベクトル積の計算

TL モデル

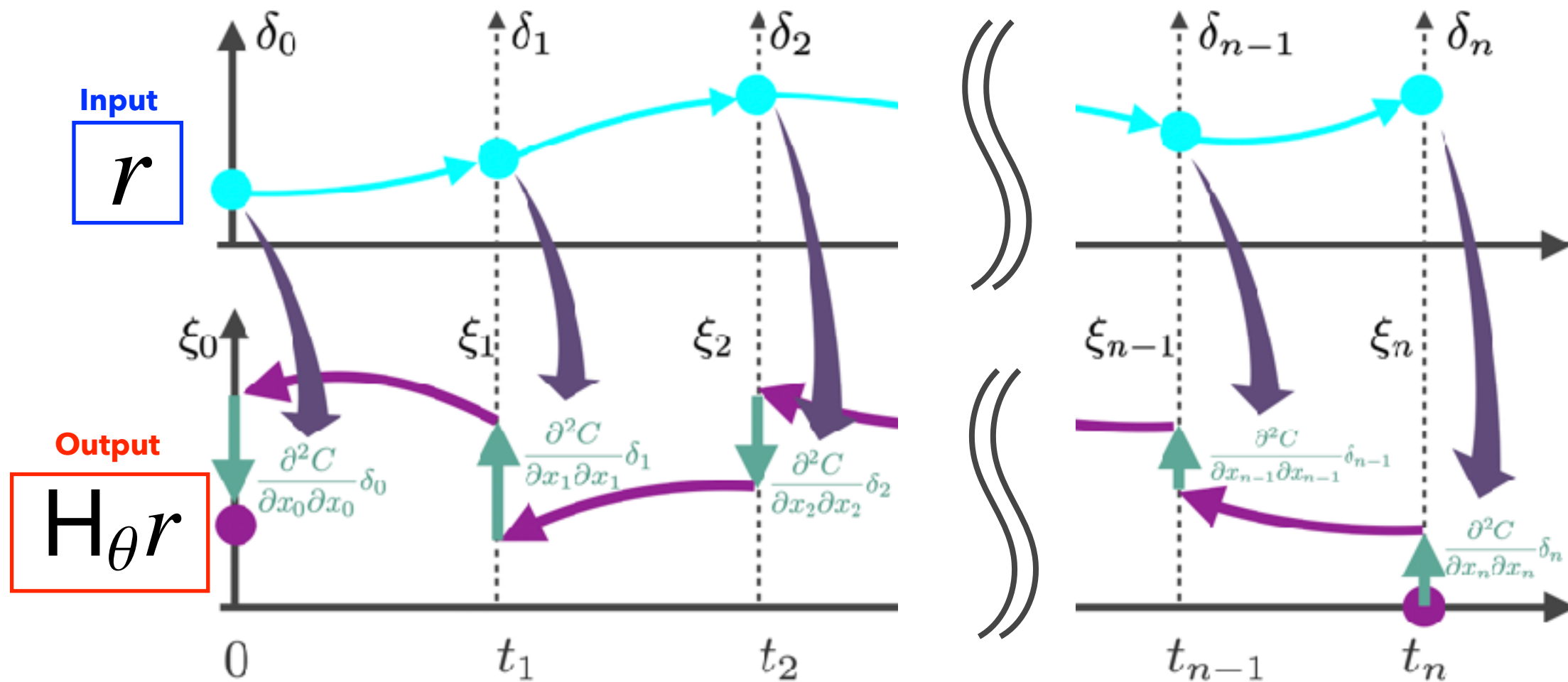
$$\frac{d}{dt}\delta_t = (\nabla_x f) \delta_t$$

$$\delta_0 = r$$

SOAモデル

$$-\frac{d}{dt}\xi_t = (\nabla_x f)^\top \xi_t + \{(\nabla_x \nabla_x f) \delta_t\}^\top \lambda_t + \sum_{t' \in T^{\text{obs}}} \delta(t - t') \nabla_{x_t} \nabla_{x_{t'}}^\top C \delta_{t'}$$

$$\xi_T = 0 \quad \xi_0 = H_\theta \delta_0 = H_\theta r$$



$$T^{\text{obs}} = \{0, t_1, t_2, \dots, t_n = T\}$$

SOA 法を用いた不確実性評価法

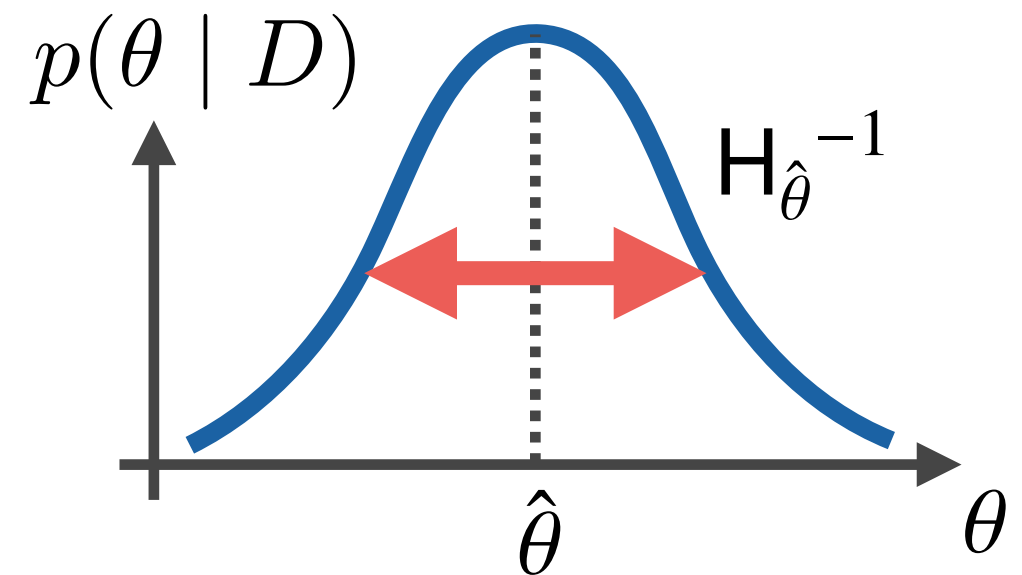
$$Hy = b$$

① y の初期値を設定

② TL モデルを解く

③ SOA モデルを時間後ろ向きに解く

④ クリロフ部分空間法で更新。②へ戻る。



$b = e_i$ に取れば $y_i = (H_{\theta}^{-1})_{i,i}$ は分散、 $y_j = (H_{\theta}^{-1})_{i,j}$ は共分散

TL, SOAモデルの計算コスト $\approx$ シミュレーションモデルの計算コスト

➡ 逆行列の計算コスト = (シミュレーション時間) $\times$ (反復回数) $\times$ (変数の数)

1 列のみの抽出であれば、(シミュレーション時間) $\times$ (反復回数)

c.f. 直接法 (シミュレーション時間) $\times$ (変数の数) 2 + (変数の数) 3

不確実性評価法の適用

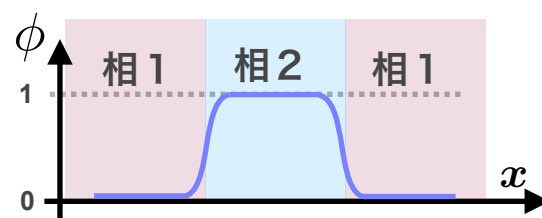
Allen-Cahn モデル (phase-field モデル)

2つの相の時空間発展を記述

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

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

$\phi(x, t)$: Phase-field 変数

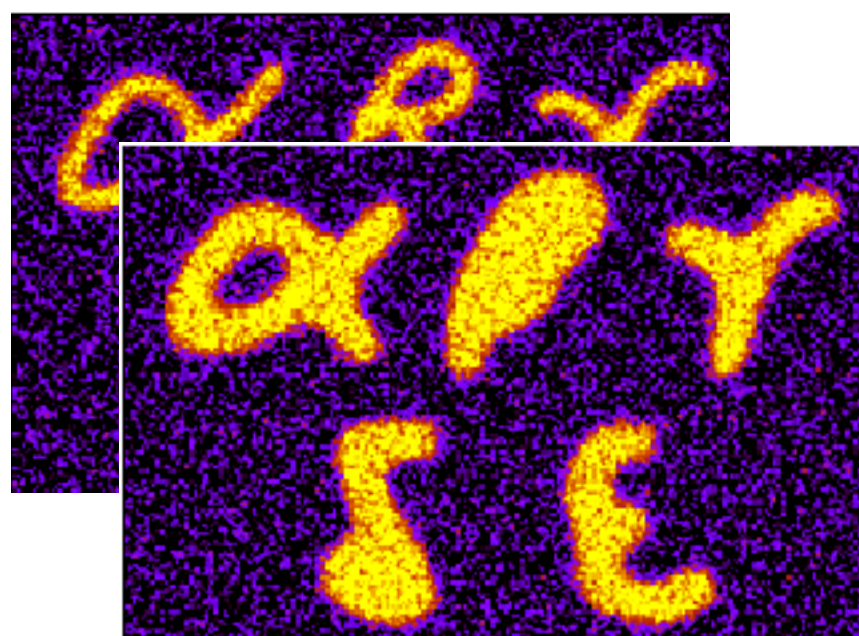


m : 相境界移動方向を決めるパラメータ

TDGL方程式、結晶成長モデル、延焼モデル、...



問題設定



Noisy なスナップショットの時系列データから

- m の不確実性 δm

(60,001次元空間中の事後分布の標準偏差 1 つだけ)

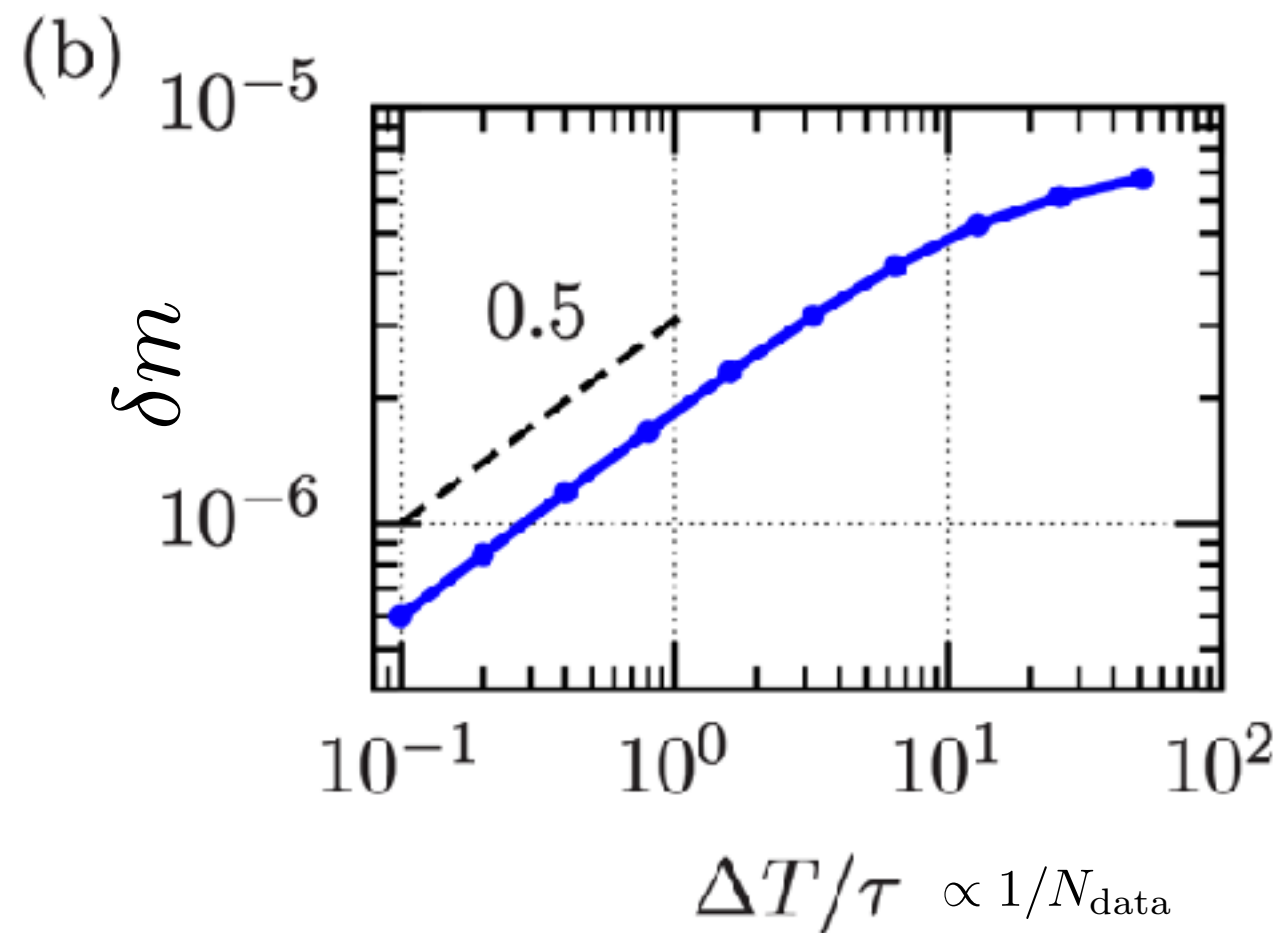
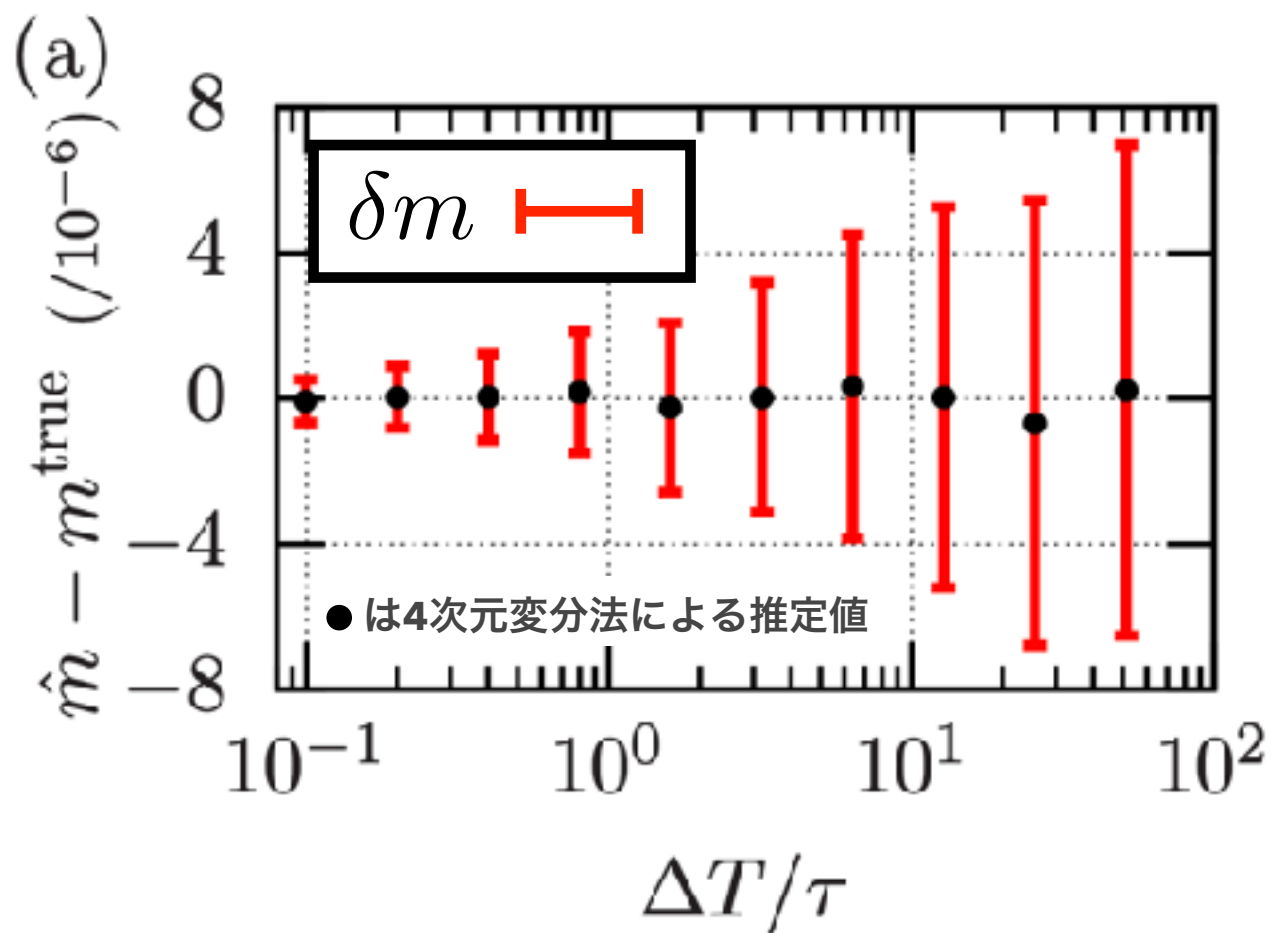
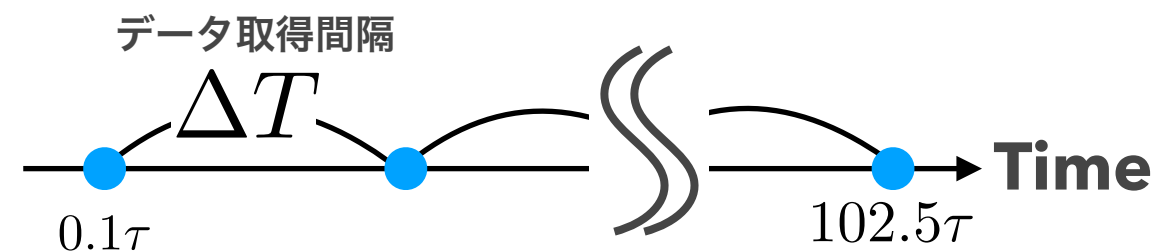
のみをSOA法によって選択的に評価

詳細な設定

事前分布: 一様分布 ノイズ分布: $q(\omega) = N(0, 0.01^2)$

データ取得: 時間窓 $[0.1\tau, 102.5\tau]$ クリロフ部分空間法: 共役残差法

不確実性評価法の適用

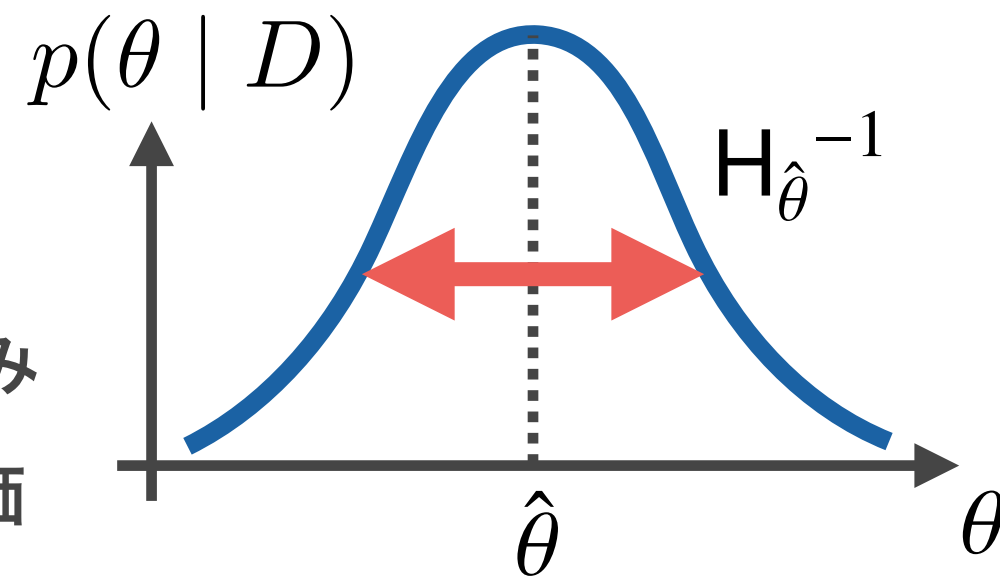


高次元の事後分布からの選択的な不確実性抽出が可能に

まとめと展望

まとめ

- 4次元変分法データ同化
- サンプルング法との相補的な高速化へ向けた試み
- **Second-order adjoint 法** に基づく不確実性評価



今後の展望

- いかに少ない試行で意味のある情報をとってこれるか
- 最適化法とサンプルング法の良いところ取り

ありがとうございました。