Coarse-grained sensitivity for multiscale data assimilation

Nozomi Sugiura*

Japan Agency for Marine-Earth Science and Technology, Yokosuka 237-0061, Japan

(Received 30 October 2015; revised manuscript received 22 February 2016; published 13 May 2016)

We show that the effective average action and its gradient are useful for solving multiscale data assimilation problems. We also present a procedure for numerically evaluating the gradient of the effective average action and demonstrate that the variational problem for slow degrees of freedom can be solved properly using the effective gradient.

DOI: 10.1103/PhysRevE.93.052212

I. INTRODUCTION

A key problem in the assimilation of data for nonlinear multiscale systems concerns the optimization of the slow degrees of freedom after the fast degrees of freedom have been properly averaged [1]. This is also the case with data assimilation for coupled atmosphere-ocean systems [2]. From a statistical point of view, this amounts to performing some integration with respect to similar realizations of control variables and packing them together into an effective cost function (action) [3] (see below for the definition). Geometrically, the rough surface of the original cost function can be smoothed according to a coarse-grained averaging procedure.

To see this, we first review how the concept of the effective action is relevant to data assimilation [4]. Data assimilation concerns the following statistical problem: Given the observation $y \in \mathbb{R}^p$, the prior probability $P[\chi]$ of the control variable $\chi \in \mathbb{R}^M$, and the likelihood $P[y|\chi]$ of the observation, the conditional expectation of any physical quantity $G[\chi]$ is calculated through the integral

$$E[G[\chi]|y] = \frac{\int d\chi\, G[\chi]P[\chi]P[y|\chi]}{\int d\chi\, P[\chi]P[y|\chi]} = \frac{\int d\chi\, G[\chi]e^{-S[\chi]}}{\int d\chi\, e^{-S[\chi]}}, \tag{1}$$

where $S[\chi]$ is called the action, or cost function, and

$$\int d\chi = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} d\chi_1 d\chi_2 \cdots d\chi_M \tag{2}$$

denotes the multiple integral over all possible combinations (paths) of χ, also called the path integral. Although the control variable $\chi = \chi(x,t)$ can generally be a field defined in some space-time (x,t), we confine ourselves to the case of a discrete space-time with M cells, that is, $\chi \in \mathbb{R}^M$. Note that Eq. (1) includes the posterior probability $P[\chi|y] = P[\chi]P[y|\chi]/P[y]$ as a special case with the δ functional $G[\chi'] = \delta[\chi' - \chi]$.

If the posterior $P[\chi|y]$ is highly concentrated around the most probable state $\hat{\chi}$, which means $P[\chi|y] \simeq \delta[\chi - \hat{\chi}]$, Eq. (1) can be approximated as

$$E[G[\chi]|y] \simeq G[\hat{\chi}]. \tag{3}$$

In this situation, it is important to find the control variable that minimizes the cost function $S[\chi]$. 4D-Var efficiently determines one of the stationary points satisfying $\delta S[\chi]/\delta\chi = 0$. However, it does not necessarily give the desired global minimum, as the cost function may have multiple minima. With regard to the shape of the cost function, we could fail to see the overall structure.

To deal with more general posterior probabilities and to calculate the conditional expectation more robustly, we present an effective alternative to the cost function. We introduce an external source term $-J^T\chi$ ($J \in \mathbb{R}^M$ is an external field) to the action in the normalization factor $\int d\chi \exp(-S[\chi])$ of Eq. (1). This leads to the following definition of the partition function:

$$Z[J] = \int d\chi\, e^{-S[\chi]+J^T\chi}, \tag{4}$$

which encodes all the information about the conditional expectation as follows:

$$E[G[\chi]|y] = \frac{1}{Z[0]} G\left[\frac{\delta}{\delta J}\right] Z[J]\bigg|_{J=0}, \tag{5}$$

where $G[\delta/\delta J]$ should be interpreted as an operator in which the argument χ of the algebraic expression $G[\chi]$ is replaced with the differential operator. The logarithm of the partition function

$$W[J] = \ln Z[J] \tag{6}$$

is also useful, because it contains all the information about the cumulants. For example,

$$\frac{\delta W}{\delta J}[0] = E[\chi^T|y], \tag{7}$$

$$\frac{\delta^2 W}{\delta J^2}[0] = E[(\chi - E[\chi|y])(\chi - E[\chi|y])^T|y]. \tag{8}$$

That is to say, we can extract information about the expected value if we perturb the external field J and observe how the normalization factor $Z[J]$, or $W[J]$, changes.

To estimate the expected value, we can construct a functional called the effective action [3,4], whose independent variable is the expected value ϕ in the presence of the external field J, through the Legendre transformation

$$\Gamma[\phi] \equiv \sup_J(-W[J] + J^T\phi). \tag{9}$$

*nsugiura@jamstec.go.jp

Taking the supremum while ϕ remains fixed in Eq. (9), we obtain

$$\phi^T = \frac{\delta W}{\delta J}[J]. \quad (10)$$

Using this and taking the derivative with respective to ϕ gives

$$\frac{\delta \Gamma}{\delta \phi}[\phi] = -\frac{\delta W}{\delta J}[J]\frac{\delta J}{\delta \phi} + \frac{\delta W}{\delta J}[J]\frac{\delta J}{\delta \phi} + J^T = J^T. \quad (11)$$

There are at least two advantages to this transformation. First, it results in a convex function because we take the Legendre transform of a convex function $W[J]$ [5]. Second, it elicits a symmetric relation between $W[J]$ and $\Gamma[\phi]$:

$$\phi^T = \frac{\delta W}{\delta J}[J], \quad J^T = \frac{\delta \Gamma}{\delta \phi}[\phi]. \quad (12)$$

This implies that its unique stationary point, which satisfies $\delta\Gamma[\phi]/\delta\phi = 0$, identifies the expected value $\phi = (\delta W[0]/\delta J)^T = E[\chi|y]$. In other words, we can find the conditional expectation by finding the stationary point of the effective action. Equation (12) also implies

$$\frac{\delta^2 \Gamma}{\delta \phi^2}[\phi] = \frac{\delta J}{\delta \phi} = \left(\frac{\delta \phi}{\delta J}\right)^{-1} = \left(\frac{\delta^2 W}{\delta J^2}[J]\right)^{-1}. \quad (13)$$

Comparing this with (8), we see that the stationary point of the effective action also provides more cumulant information.

Since $\Gamma[\phi]$ in Eq. (9) should be regarded as a function of ϕ alone, we eliminate J using Eq. (12) to obtain

$$\Gamma[\phi] = -W\left[\frac{\delta \Gamma}{\delta \phi}[\phi]\right] + \frac{\delta \Gamma}{\delta \phi}[\phi]\phi \quad (14)$$

$$= -\ln\left\{\int d\chi\, e^{-S[\chi]+(\delta\Gamma[\phi]/\delta\phi)\chi}\right\} + \frac{\delta \Gamma}{\delta \phi}[\phi]\phi \quad (15)$$

$$= -\ln\left\{\int d\chi\, e^{-S[\chi]+(\delta\Gamma[\phi]/\delta\phi)(\chi-\phi)}\right\}. \quad (16)$$

This suggests a means of calculating the effective action. However, it requires a recursive procedure that includes integrations over all possible combinations of control variables χ, which appears to be intractable.

To compute the effective action, we may evaluate this integral stepwise, using methods developed in renormalization group theory [6]. A relevant concept that we will explore later is the effective average action $\Gamma_k[\phi]$ proposed by Wetterich [7], which constitutes a one-parameter family of functionals interpolating between the action $S[\phi]$ and the effective action $\Gamma[\phi]$.

The aim of this paper is to propose a possible framework that will help solve the multiscale data assimilation problem by replacing the cost function with the effective average action. We also propose a method for evaluating the gradient of the effective average action numerically. In principle, this enables us to solve a broader range of data assimilation problems by seeking the stationary point of the effective average action using its gradient, which is referred to as the effective gradient or the coarse-grained sensitivity.

The concept of the effective average action is explained in Sec. II A and the meaning of its stationary point is clarified in Sec. II B. Section III describes a procedure for calculating the gradient of the effective action. Sections IV and V illustrate some applications of the method to data assimilation or sensitivity studies.

II. EFFECTIVE AVERAGE ACTION

When dealing with a multiscale system, it is often difficult to define the sensitivity with respect to the control variable, because fast degrees of freedom may have many statistical paths related to the sensitivity that cannot be expressed in a deterministic manner. In other words, we cannot always use the sensitivity to choose the optimal realization of a fluctuation from among a large ensemble of fast fluctuations in the control space. This motivates the definition of a macroscopic field in which fast degrees of freedom are treated as averaged quantities. A suitable tool for this purpose is the effective average action [7].

A. Definition

The procedure for the effective action reviewed in the Introduction, $S[\chi] \rightarrow W[J] \rightarrow \Gamma[\phi]$, can also be applied to the derivation of the effective average action. This introduces some filtering terms, $\Delta S_k[\chi]$ to Eq. (4) and $-\Delta S_k[\phi]$ to Eq. (9), which have the effect of selectively integrating out the fast degrees of freedom in the control space to enable the dynamics of slower variables to be investigated.

We start by defining an infrared filter [8]

$$\Delta S_k[\chi] \equiv \tfrac{1}{2}\chi^T R_k \chi, \quad (17)$$

where $\chi \in \mathbb{R}^M$ is the control variable in a discrete space-time with M cells and $R_k \in \mathbb{R}^M \times \mathbb{R}^M$ is a discrete low-pass filter. To derive concrete expressions for R_k, let us consider a simple case with a cyclic control variable χ in a one-dimensional discrete domain $l = 1,2,\ldots,M$. We define the discrete Fourier transform $\hat{\chi}$ of χ and its inverse as

$$\hat{\chi}_j = \frac{1}{\sqrt{M}}\sum_{l=1}^{M} \chi_l e^{-(2\pi jl/M)i}, \quad |j| < [M/2] \quad (18)$$

$$\chi_l = \frac{1}{\sqrt{M}}\sum_{j=-[M/2]}^{[M/2]} \hat{\chi}_j e^{(2\pi jl/M)i}, \quad l = 1,2,\ldots,M, \quad (19)$$

where M is odd for simplicity, $[\cdot]$ denotes the roundoff, and i is the imaginary unit. If we assume that the infrared filter is represented by a cutoff of high-wave-number modes, then

$$\Delta S_k[\hat{\chi}] = \frac{1}{2}\sum_j \hat{\chi}_{-j}\hat{R}_k(j)\hat{\chi}_j, \quad (20)$$

$$\hat{R}_k(j) = \begin{cases} k^2\left(1 - \frac{j^2}{j_k^2}\right) & \text{if } |j| < j_k \\ 0 & \text{otherwise,} \end{cases} \quad (21)$$

where $j, j_k \in \mathbb{Z}$ and j_k is the cutoff level. The filtering term can then be written as

$$\Delta S_k[\hat{\chi}] = \frac{1}{2}\sum_{|j|<j_k} \hat{\chi}_{-j} k^2 \left(1 - \frac{j^2}{j_k^2}\right)\hat{\chi}_j$$

$$= \frac{1}{2}\sum_{l,l'=1}^{M} \chi_l \left\{\sum_{|j|<j_k} \frac{k^2}{M}\left(1 - \frac{j^2}{j_k^2}\right) e^{[-2\pi j/(l-l')M]i}\right\}\chi_{l'} \quad (22)$$

$$= \frac{1}{2} \sum_{l,l'=1}^{M} \chi_l \left\{ \sum_{|j|<j_k} \frac{k^2}{M} \left(1 - \frac{j^2}{j_k^2}\right) \times \cos\left[\frac{2\pi j(l-l')}{M}\right] \right\} \chi_{l'}. \quad (23)$$

The expression in curly brackets gives a matrix representation of R_k in position space. If M is even, we can replace Eq. (21) by

$$\hat{R}_k\left(j + \frac{1}{2}\right) = \begin{cases} k^2\left(1 - \frac{(j+\frac{1}{2})^2}{j_k^2}\right) & \text{if } |j + \frac{1}{2}| < j_k \\ 0 & \text{otherwise,} \end{cases} \quad (24)$$

where $j, j_k \in \mathbb{Z}$. To further simplify the filter, we can also use $R_k = k^2$ with $j_k \to \infty$, which yields

$$\Delta S_k[\chi] = \frac{k^2}{2} \sum_{l=1}^{M} \chi_l^2. \quad (25)$$

As we will see later, the filter should have the following properties:

$$k \to \infty \Rightarrow R_k \to \infty, \quad (26)$$

$$k \to 0 \Rightarrow R_k \to 0. \quad (27)$$

With the filtering term, the partition function Z_k and its logarithm for the high-wave-number modes can be defined as

$$Z_k[J] = \int d\chi\, e^{-S[\chi] - \Delta S_k[\chi] + J^T \chi}, \quad (28)$$

$$W_k[J] = \ln Z_k[J], \quad (29)$$

where J is the external field. As $\Delta S_k[\chi]$ is large for low-wave-number modes, the term $\exp(-\Delta S_k[\chi])$ has the effect of focusing the integration on the high-wave-number modes in χ.

Applying the Legendre transformation to switch the independent variable J to ϕ, we obtain the effective average action [7]

$$\Gamma_k[\phi] \equiv \sup_J (-W_k[J] + J^T \phi) - \Delta S_k[\phi]. \quad (30)$$

Note that this transform should have the additional term $-\Delta S_k[\phi]$ for the following reason [7]. From the supremum condition, we find

$$\phi = \left(\frac{\delta W_k}{\delta J}[J]\right)^T = \langle \chi \rangle_{k,J} \equiv \frac{\int d\chi\, \chi\, e^{-S[\chi] - \Delta S_k[\chi] + J^T \chi}}{\int d\chi\, e^{-S[\chi] - \Delta S_k[\chi] + J^T \chi}} \quad (31)$$

$$= \frac{\int d\chi (\phi + \chi) e^{-S[\phi+\chi] - \Delta S_k[\chi] + (J^T - \phi^T R_k)\chi}}{\int d\chi\, e^{-S[\phi+\chi] - \Delta S_k[\chi] + (J^T - \phi^T R_k)\chi}}. \quad (32)$$

Equation (31) appears to indicate that ϕ is analogous to the conditional expectation of the control variable in the vicinity of $\chi = 0$ under the existence of the external field J. However, if we insert into Eq. (32) the relation

$$\frac{\delta \Gamma_k}{\delta \phi}[\phi] = J^T - \frac{\delta \Delta S_k}{\delta \phi}[\phi], \quad (33)$$

which is derived by the same operation as in Eq. (11), we obtain

$$\phi = \frac{\int d\chi (\phi + \chi) e^{-S[\phi+\chi] - \Delta S_k[\chi] + (\delta \Gamma_k[\phi]/\delta\phi)\chi}}{\int d\chi\, e^{-S[\phi+\chi] - \Delta S_k[\chi] + (\delta \Gamma_k[\phi]/\delta\phi)\chi}}. \quad (34)$$

This shows that ϕ is in fact analogous to the conditional expectation of the control variable in the vicinity of ϕ itself under the existence of the external field $\delta\Gamma_k[\phi]/\delta\phi$. Thus, the term $-\Delta S_k[\phi]$ in Eq. (30) ensures that ϕ is always the average of the surrounding χ.

From Eq. (31) and the derivative of Eq. (33) with respect to ϕ, we can see that the effective average action $\Gamma_k[\phi]$ also satisfies the following equality:

$$\frac{\delta^2 \Gamma_k}{\delta \phi^2}[\phi] = \frac{\delta J}{\delta \phi} - \frac{\delta^2 \Delta S_k}{\delta \phi^2} \quad (35)$$

$$= \left(\frac{\delta^2 W_k}{\delta J^2}[J]\right)^{-1} - \frac{\delta^2 \Delta S_k}{\delta \phi^2} \quad (36)$$

$$= (\langle \chi \chi^T \rangle_{k,J} - \phi \phi^T)^{-1} - \frac{\delta^2 \Delta S_k}{\delta \phi^2}. \quad (37)$$

Eliminating J from Eq. (30) using Eq. (33), we have

$$\Gamma_k[\phi] = -W\left[\frac{\delta \Gamma_k}{\delta \phi}[\phi] + \frac{\delta \Delta S_k}{\delta \phi}[\phi]\right] + \left(\frac{\delta \Gamma_k}{\delta \phi}[\phi] + \frac{\delta \Delta S_k}{\delta \phi}[\phi]\right)\phi - \Delta S_k[\phi] \quad (38)$$

$$= -\ln\left\{\int d\chi\, e^{-S[\chi] + (\delta\Gamma_k[\phi]/\delta\phi)(\chi - \phi) - \Delta S_k[\chi - \phi]}\right\} \quad (39)$$

$$= -\ln\left\{\int d\chi\, e^{-S[\phi+\chi] + (\delta\Gamma_k[\phi]/\delta\phi)\chi - \Delta S_k[\chi]}\right\}. \quad (40)$$

Although this has a recursive form about $\Gamma_k[\phi]$, we can write an approximation in closed form (see Appendix A 1 for the derivation):

$$\Gamma_k[\phi] \simeq S[\phi] + \frac{1}{2} \ln\det\left[\frac{\delta^2}{\delta\phi^2}(S + \Delta S_k)[\phi]\right]. \quad (41)$$

Note that we will not resort to such perturbation expansions in our numerical calculation, because it requires higher derivatives of the action, which are not always easy to calculate.

Figure 1 illustrates the relation between the action and the effective average action in a simple case, where we assume $\chi \in \mathbb{R}^1$ and the filter is $R_k = k^2$. The tangent point $A(\phi, \Gamma_k[\phi])$ and the slope $\delta\Gamma_k[\phi]/\delta\phi$ are such that the point ϕ coincides with the weighted average of the points χ around the interval $\phi - k^{-1} \leqslant \chi \leqslant \phi + k^{-1}$. Indeed, we see from Eqs. (34) and (40) that

$$\phi = \int d\chi\, \chi\, e^{-S[\chi] + \Gamma_k[\phi] + (\delta\Gamma_k[\phi]/\delta\phi)(\chi - \phi) - (k^2/2)(\chi - \phi)^2}. \quad (42)$$

One interesting thing about this smoothing is that it is not an averaging of the value of $S[\chi]$ but of its independent variable χ. Thereby, the effective average action serves as a kind of smoothed version of the cost function.

Taking Eqs. (26) and (27) into account, $k \to \infty$ implies $e^{-\Delta S_k[\chi]} = e^{-\chi^T R_k \chi/2} \to \delta[\chi]$, which leads Eq. (40)

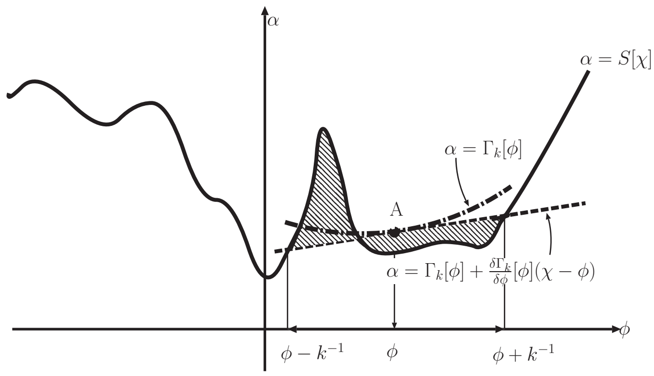

FIG. 1. Concept of the effective average action Γ_k (dot-dashed curve). The tangent point $A(\phi,\Gamma_k[\phi])$ and the slope $\delta\Gamma_k[\phi]/\delta\phi$ are such that the point ϕ coincides with the weighted average of the points χ around the interval $\phi - k^{-1} \leqslant \chi \leqslant \phi + k^{-1}$. The weight exponentiates the deviation (shaded region) of the action $\alpha = S[\chi]$ (solid curve) from the tangent plane $\alpha = \Gamma_k[\phi] + (\delta\Gamma_k[\phi]/\delta\phi)(\chi - \phi)$ (dashed line). For simplicity, it is assumed that χ is one dimensional and the filter has $R_k = k^2$.

to

$$\Gamma_{k\to\infty}[\phi] = S[\phi]. \tag{43}$$

This means that when k is sufficiently large no fields around ϕ are counted in $\Gamma_k[\phi]$ other than the field ϕ itself and thus the effective average action approaches the original action, or the cost function. In contrast, it is apparent that

$$\Gamma_{k\to 0}[\phi] = \Gamma[\phi]. \tag{44}$$

Hence, we have confirmed that $\Gamma_k[\phi]$ constitutes a one-parameter family of functionals interpolating between the action $S[\phi]$ and the effective action $\Gamma[\phi]$. Hereafter, we assume that k has a finite value so as to integrate out some modes.

B. Property of the stationary point

We assume that the stationary problem

$$\frac{\delta\Gamma_k}{\delta\phi}[\phi] = J^T - \frac{\delta\Delta S_k}{\delta\phi}[\phi] = 0 \tag{45}$$

has stationary values at $\hat{\phi}$. From the definition (31), we have

$$\hat{\phi} = \langle\chi\rangle_{k,(\delta\Delta S_k/\delta\phi)[\hat{\phi}]}. \tag{46}$$

Since Eq. (45) can be thought of as the statistical equation of motion for the field ϕ, the solution $\hat{\phi}$ offers the estimated path for the statistical problem. At the stationary point, Eq. (32) reads

$$\hat{\phi} = \frac{\int d\chi(\hat{\phi} + \chi)e^{-S[\hat{\phi}+\chi]-\Delta S_k[\chi]}}{\int d\chi\, e^{-S[\hat{\phi}+\chi]-\Delta S_k[\chi]}}. \tag{47}$$

This means that the stationary value $\hat{\phi}$ provides the average with respect to high-wave-number modes (see also Appendix A 2).

III. ESTIMATION OF THE EFFECTIVE GRADIENT

A. Definition as an expected value

Rewriting Eq. (40), we introduce the exponent $\mathfrak{R}[\phi,\chi]$ for convenience:

$$e^{-\Gamma_k[\phi]} = e^{-S[\phi]}\int d\chi\, e^{-\mathfrak{R}[\phi,\chi]}, \tag{48}$$

$$\mathfrak{R}[\phi,\chi] \equiv S[\phi+\chi] - S[\phi] - \frac{\delta\Gamma_k}{\delta\phi}[\phi]\chi + \Delta S_k[\chi], \tag{49}$$

$$\Gamma_k[\phi] = S[\phi] - \ln\int d\chi\, e^{-\mathfrak{R}[\phi,\chi]}. \tag{50}$$

The gradient of $\Gamma_k[\phi]$ is derived as the expected value $\langle\cdot\rangle_{\mathfrak{R}}$ under the weight $e^{-\mathfrak{R}}$:

$$\begin{aligned}\frac{\delta\Gamma_k}{\delta\phi}[\phi] &= \frac{\delta S}{\delta\phi}[\phi] - \frac{\int d\chi\left(-\frac{\delta\mathfrak{R}}{\delta\phi}[\phi,\chi]\right)e^{-\mathfrak{R}[\phi,\chi]}}{\int d\chi\, e^{-\mathfrak{R}[\phi,\chi]}}\\ &= \left\langle\frac{\delta S}{\delta\phi}[\phi+\chi]\right\rangle_{\mathfrak{R}} - \langle\chi\rangle_{\mathfrak{R}}^T\frac{\delta^2\Gamma_k}{\delta\phi^2}[\phi].\end{aligned} \tag{51}$$

From Eq. (34) we have that

$$0 = \int d\chi\,\chi\, e^{-\mathfrak{R}[\phi,\chi]} \propto \langle\chi\rangle_{\mathfrak{R}}. \tag{52}$$

Thus, Eq. (51) can be simplified to

$$\frac{\delta\Gamma_k}{\delta\phi}[\phi] = \left\langle\frac{\delta S}{\delta\phi}[\phi+\chi]\right\rangle_{\mathfrak{R}}. \tag{53}$$

Equations (49) and (53) are recursive with respect to $\delta\Gamma_k[\phi]/\delta\phi$. Therefore, we need some approximation to enable a numerical evaluation. We may replace $\delta\Gamma_k[\phi]/\delta\phi$ with the approximation $\delta S[\phi]/\delta\phi$ in $\mathfrak{R}[\phi,\chi]$ on the right-hand side of the equation. Equation (53) can then be evaluated using the Metropolis method [9,10]. We can then apply a successive correction procedure by updating the expectation with the latest value of $\delta\Gamma_k[\phi]/\delta\phi$ in the weight. Furthermore, from Eqs. (31), (34), and (37) we find that the second derivative can also be derived as the expected value

$$\frac{\delta^2\Gamma_k}{\delta\phi^2}[\phi] = \langle\chi\chi^T\rangle_{\mathfrak{R}}^{-1} - \frac{\delta^2\Delta S_k}{\delta\phi^2}. \tag{54}$$

B. Evaluation through the Metropolis method

In contrast to the case of the effective action in Eq. (16), the filtering term $\Delta S_k[\chi]$ in Eq. (49) has the effect of confining the weight $\exp(-\mathfrak{R}[\phi,\chi])$ to a small region in the control space, as illustrated in Fig. 1. Owing to this, we may assume that the expected value will be efficiently evaluated by a Markov-chain Monte Carlo method, e.g., the Metropolis adjusted Langevin algorithm [11]. Using the fact that the Langevin equation

$$d\chi_t = \tfrac{1}{2}\nabla\ln f(\chi_t)dt + dW_t \tag{55}$$

(where W_t is the Wiener process) has the invariant distribution $\pi(\chi) \equiv f(\chi)/\int d\chi\, f(\chi)$, we construct a Markov chain by discretizing the equation and applying an acceptance and rejection procedure.

At time step n, according to the weight

$$f(\chi^{(n)}) = e^{-\mathfrak{R}[\phi,\chi^{(n)}]}, \tag{56}$$

we define a proposal normal distribution

$$q(\chi|\chi^{(n)}) = \mathcal{N}\left(\chi^{(n)} - \frac{\sigma^2}{2}\nabla_{\chi^{(n)}}\mathfrak{R}[\phi,\chi^{(n)}],\sigma^2 I\right), \quad (57)$$

$$\nabla_{\chi^{(n)}}\mathfrak{R}[\phi,\chi^{(n)}] = \left(\frac{\delta\mathfrak{R}}{\delta\chi^{(n)}}[\phi,\chi^{(n)}]\right)^T, \quad (58)$$

$$\frac{\delta\mathfrak{R}}{\delta\chi^{(n)}}[\phi,\chi^{(n)}] = \frac{\delta S}{\delta\phi}[\phi+\chi^{(n)}] - \frac{\delta\Gamma_k}{\delta\phi}[\phi] + \frac{\delta\Delta S_k}{\delta\phi}[\chi^{(n)}] \quad (59)$$

$$\simeq \frac{\delta S}{\delta\phi}[\phi+\chi^{(n)}] - \frac{\delta S}{\delta\phi}[\phi] + \frac{\delta\Delta S_k}{\delta\phi}[\chi^{(n)}] \quad (60)$$

to generate a random field χ^* that obeys q:

$$\chi^* = \chi^{(n)} - \frac{\sigma^2}{2}\nabla_{\chi^{(n)}}\mathfrak{R}[\phi,\chi^{(n)}] + \sigma\xi, \quad \xi \sim \mathcal{N}(0,I). \quad (61)$$

We then update $\chi^{(n+1)} = \chi^*$ with the acceptance probability

$$\rho(\chi^{(n)},\chi^*) = \min\left(1, \frac{f(\chi^*)}{f(\chi^{(n)})}\frac{q(\chi^{(n)}|\chi^*)}{q(\chi^*|\chi^{(n)})}\right), \quad (62)$$

or retain $\chi^{(n+1)} = \chi^{(n)}$. The ensemble of sample sequences $\chi^{(n)}$ drawn in this way approximately follows the invariant distribution $\pi(\chi)$ and we can estimate the expected value accordingly:

$$\left\langle\frac{\delta S}{\delta\phi}[\phi+\chi]\right\rangle_{\mathfrak{R}} \simeq \frac{1}{N}\sum_{n=1}^{N}\frac{\delta S}{\delta\phi}[\phi+\chi^{(n)}]. \quad (63)$$

Note that, after averaging, we may perform another refined averaging by substituting the derived sensitivity into the weight, because the term $(\delta S[\phi]/\delta\phi)^T$ in Eq. (60) should have been $(\delta\Gamma_k[\phi]/\delta\phi)^T$. Furthermore, within the limit of the accuracy of importance sampling, the finite difference of Γ_k can also be estimated as

$$\Gamma_k[\phi+\Delta\phi] - \Gamma_k[\phi] = -\ln\langle e^{-\mathfrak{R}[\phi+\Delta\phi,\chi]+\mathfrak{R}[\phi,\chi]}\rangle_{\mathfrak{R}}. \quad (64)$$

IV. SIMPLE EXAMPLE

As a simple example, we consider a double-well potential

$$S[\phi] = \tfrac{1}{2}(\phi^2 - a^2)^2, \quad a = 0.5, \quad R_k = 2^2. \quad (65)$$

The effective average action given by a perturbation expansion up to the second order (see Appendix A 1) is

$$\Gamma[\phi] = S[\phi] + \tfrac{1}{2}\ln(6\phi^2 - 2a^2 + R_k). \quad (66)$$

Figure 2 shows that the potential barrier at the center (black curve) is eliminated (purple curve) by integrating out the fluctuation. The gradients, along with the effective gradient of the action evaluated by the Metropolis method, are shown in Fig. 3. These curves suggest that the variational method using the coarse-grained sensitivity can capture the expected value $\hat{\phi} = 0$, which traditional variational methods will fail to find. This is because, in principle, variational methods are all designed to find one of the stationary values of the cost function, in this case $\phi = \pm a$.

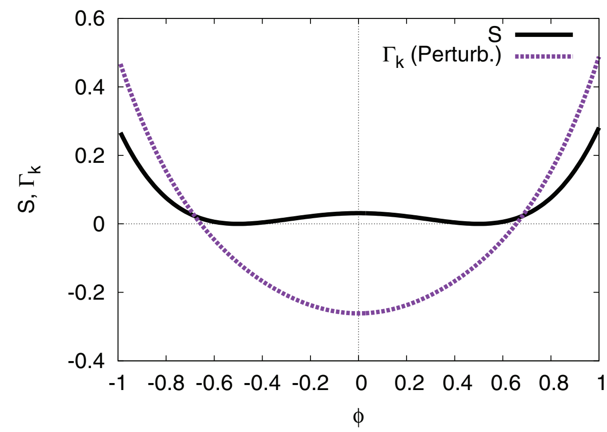

FIG. 2. Action $S[\phi]$ (black curve) and effective average action $\Gamma_k[\phi]$ evaluated using a perturbation expansion (purple curve) for the double-well potential.

V. APPLICATION TO DATA ASSIMILATION

A. Quadratic cost function

The cost function used in data assimilation is usually a quadratic in the nonlinear functional $F[\phi]$, such as [12]

$$S[\phi] \equiv \frac{1}{2}F[\phi]^T F[\phi], \quad \frac{\delta S}{\delta\phi}[\phi] = F[\phi]^T\frac{\delta F}{\delta\phi}[\phi]. \quad (67)$$

To obtain the averaged sensitivity $\delta\Gamma_k[\phi]/\delta\phi = \langle\delta S[\phi+\chi]/\delta\phi\rangle_{\mathfrak{R}}$, we perform the following calculations:

$$\left(\frac{\delta S}{\delta\phi}[\phi+\chi]\right)^T = \left(\frac{\delta F}{\delta\phi}[\phi+\chi]\right)^T F[\phi+\chi], \quad (68)$$

$$\mathfrak{R}[\phi,\chi] = \frac{1}{2}F[\phi+\chi]^T F[\phi+\chi] - \frac{1}{2}F[\phi]^T F[\phi] - \chi^T\left(\frac{\delta\Gamma_k}{\delta\phi}[\phi]\right)^T + \Delta S_k[\chi] \quad (69)$$

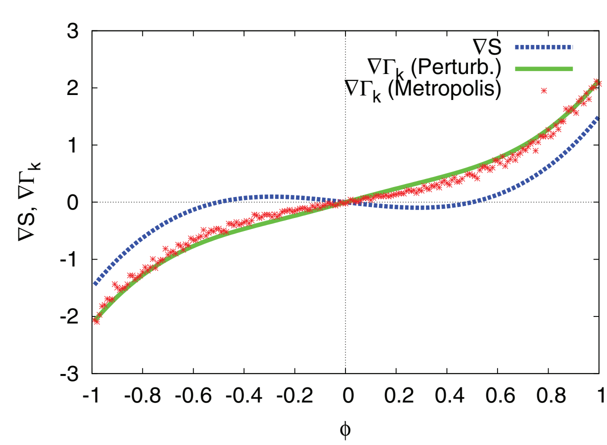

FIG. 3. Gradient of action $\delta S[\phi]/\delta\phi$ (blue curve), effective gradient of action $\delta\Gamma_k[\phi]/\delta\phi$ evaluated using a perturbation expansion (green curve), and effective gradient of action evaluated by the Metropolis method (red curve) for the double-well potential.

$$\simeq \frac{1}{2}F[\phi+\chi]^T F[\phi+\chi] - \frac{1}{2}F[\phi]^T F[\phi] - \chi^T\left(\frac{\delta F}{\delta\phi}[\phi]\right)^T F[\phi] + \Delta S_k[\chi], \quad (70)$$

$$\nabla_\chi \mathfrak{R}[\phi,\chi] = \left(\frac{\delta F}{\delta\phi}[\phi+\chi]\right)^T F[\phi+\chi] - \left(\frac{\delta\Gamma_k}{\delta\phi}[\phi]\right)^T + \left(\frac{\delta\Delta S_k}{\delta\chi}[\chi]\right)^T. \quad (71)$$

Thus, each sample requires a forward integration $F[\phi+\chi]$ and a subsequent adjoint integration $(\delta F[\phi+\chi]/\delta\phi)^T$.

B. Logistic map

We now consider a smoothing problem for the logistic map $\Psi \mapsto r\Psi(1-\Psi)$ [13,14] fitted to observation y. As a data assimilation problem, we use the following cost function S and its gradient [15]:

$$S[\phi] = \frac{1}{2\sigma_0^2}(\phi - m_0)^2 + \sum_{j=0}^{J-1}\frac{1}{2\gamma^2}(y_{j+1} - \Psi^{(j+1)}[\phi])^2, \quad (72)$$

$$\Psi^{(j+1)}[\phi] = r\Psi^{(j)}[\phi](1-\Psi^{(j)}[\phi]), \quad \Psi^{(0)}[\phi] = \phi, \quad (73)$$

$$\frac{\delta S}{\delta\phi}[\phi] = \frac{1}{\sigma_0^2}(\phi - m_0) + \frac{1}{\gamma^2}\sum_{j=0}^{J-1}\left[\prod_{l=0}^{j} r(1-2\Psi^{(l)}[\phi])\right] \times (\Psi^{(j+1)}[\phi] - y_{j+1}), \quad (74)$$

where σ_0^2 and γ^2 are the background and observational error variances, respectively. The parameters are set to $r = 4$, $J = 6$, $\sigma_0 = 0.1$, $\gamma = 0.2$, and $m_0 = 0.4$ (first guess), similar to those in [14]. The observations are sampled from a model sequence given by the initial value $v_0 = 0.3$ added to observational noise.

The optimization problem can apparently be solved using a variational method that seeks the optimal initial condition using the gradient information. However, there are multiple extrema of the cost function (see the black curve in Fig. 4), which makes it difficult to find the global minimum. Thus, we should utilize the effective gradient (63), derived using the Metropolis method. We use a finite constant $R_k = (0.008)^{-2}$, where 0.008 is the typical half wavelength of short fluctuations in $S[\phi]$.

Figure 4 shows the action (black curve), the gradient of action (blue curve), and the effective gradient of action (red curve) for this system. It is clear that the original gradient has too many zeros for worthwhile variational data assimilation using the gradient. However, because the effective gradient has relatively few zeros, it can be applied to a variational data assimilation to find the minimum at around $\phi = v_0$, as long as the first guess is not far from the true value.

We performed two data assimilation experiments using the steepest descent method with the gradient $\delta S[\phi]/\delta\phi$ and the effective gradient $\delta\Gamma_k[\phi]/\delta\phi$ (see the algorithm in Appendix B). As shown in Fig. 5, the case with $\delta S[\phi]/\delta\phi$ converges to a local minimum $\phi = 0.388$, whereas the case with $\delta\Gamma_k[\phi]/\delta\phi$ converges to $\phi = 0.315$ around the global minimum, which indicates the superiority of the effective gradient. Note that the decrease in the cost function $S[\phi]$ in the latter case is not necessarily monotonic, because the optimization problem is actually defined for the effective average action $\Gamma_k[\phi]$. This illustrates the potential usefulness of the effective gradient (sensitivity) for data assimilation.

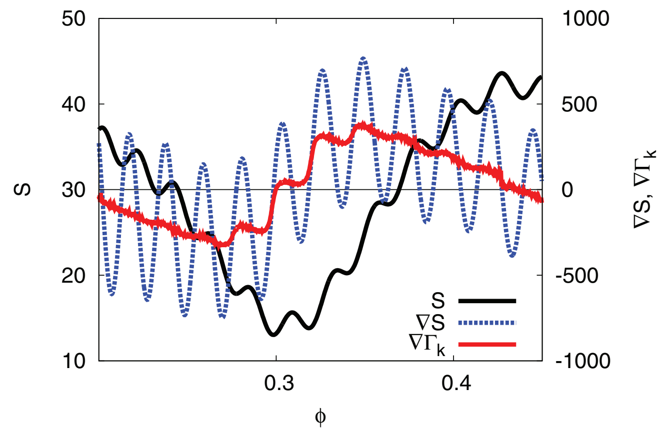

FIG. 4. Action $S[\phi]$ (black curve), gradient of action $\delta S[\phi]/\delta\phi$ (blue curve), and effective gradient of action $\delta\Gamma_k[\phi]/\delta\phi$ (red curve) for the logistic map. The true value for the data assimilation problem is $\phi = 0.3$ and the first guess is 0.4.

C. Lorenz model

Next we examine the sensitivity that appears during data assimilation in the Lorenz model, which is a simple dynamical system designed to mimic the dynamics of Rossby waves in atmospheric dynamics [14,16]. This system can be written as

$$\frac{d\theta_l}{dt} = \theta_{l-1}(\theta_{l+1} - \theta_{l-2}) - \theta_l + F, \quad l = 1,2,\ldots,M \quad (75)$$

$$\theta_0 = \theta_M, \quad \theta_{M+1} = \theta_1, \quad \theta_{-1} = \theta_{M-1}. \quad (76)$$

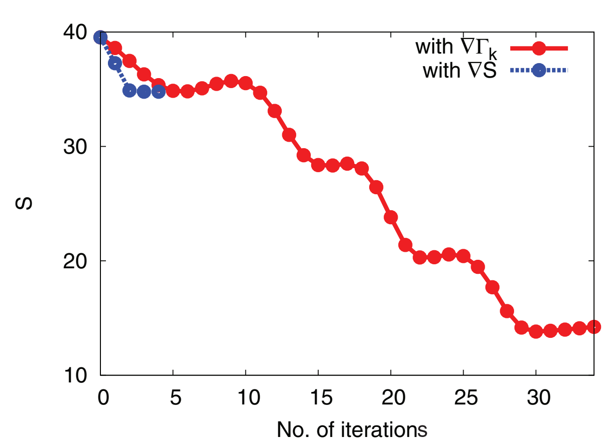

FIG. 5. Variation of the cost function during data assimilation using the steepest descent method with $\delta S[\phi]/\delta\phi$ (blue curve) and $\delta\Gamma_k[\phi]/\delta\phi$ (red curve). The former converges to a local minimum $\phi = 0.388$; the latter converges to $\phi = 0.315$, which is close to the global minimum.

We write the time evolution operator from the initial condition as

$$\Psi^{(j)}[\theta(t=0)] = \theta(t=t_j), \tag{77}$$

with time step $t_{j+1} - t_j = \Delta t$. If we choose the initial condition $\theta(t=0) = \phi \in \mathbb{R}^M$ as the control variable, we can define a cost function similar to that in Eq. (72):

$$S[\phi] = \frac{1}{2\sigma_0^2}\|\phi - m_0\|^2 + \sum_{j=0}^{J-1} \frac{1}{2\gamma^2}\|y_{j+1} - \Psi^{(j+1)}[\phi]\|^2, \tag{78}$$

where $\|\cdot\|$ is the Euclidean norm. The parameters are set to $F = 6$, $J = 32$, $M = 20$, $\sigma_0 = 2$, and $\gamma = 2$. With these parameters, this model is in an unstable regime with the first Lyapunov exponent $\lambda_1 \simeq 0.84 > 0$. The true initial condition $v_0 \in \mathbb{R}^M$ is given by model integration within an interval from a randomly chosen initial condition. The observation is sampled from a model sequence starting from v_0, with observational noise added to the sample. The observation is defined only at the times $j+1 = 3,6,\ldots,30$ and space $l = 1,2,\ldots,8$. The first guess m_0 is given by changing only the first component of v_0:

$$(m_0)_1 = (v_0)_1 + \sigma_0, \quad (m_0)_2 = (v_0)_2, \ldots, \quad (m_0)_M = (v_0)_M. \tag{79}$$

The filter R_k is in the form of Eq. (24) with $j_k = 10$ and $k = 0.25^{-1}$. Here k is set so that the filtering term $\Delta S_k[\phi]$ is of order 1, that is, $O(1) = \Delta S_k[\phi] \simeq k^2 \mathrm{var}(\phi)$, where the typical fluctuation is assumed to be $\mathrm{var}(\phi) \simeq k^{-2} = 0.25^2$. The time evolution of (75) and its adjoint are solved by the Runge-Kutta method with time step $\Delta t = 0.1$. The experiments are designed to investigate how the action, the gradient of action, and the effective gradient of action change if we move the control variable as

$$\phi_1 = (v_0)_1 + \sigma_0\eta, \quad \phi_2 = (v_0)_2, \ldots, \quad \phi_M = (v_0)_M, \tag{80}$$

where $-1 \leqslant \eta \leqslant 1$.

Figure 6 shows the action (black curve), the gradient of action (blue curve), and the effective gradient of action (red curve) for this experiment. It is clear that the original gradient has several zeros that will complicate the variational data assimilation using the gradient. However, the effective gradient has only one zero near the true value, which is a similar result as for the logistic map. Thus, this example also shows the potential usefulness of the effective gradient for finding the optimal initial condition.

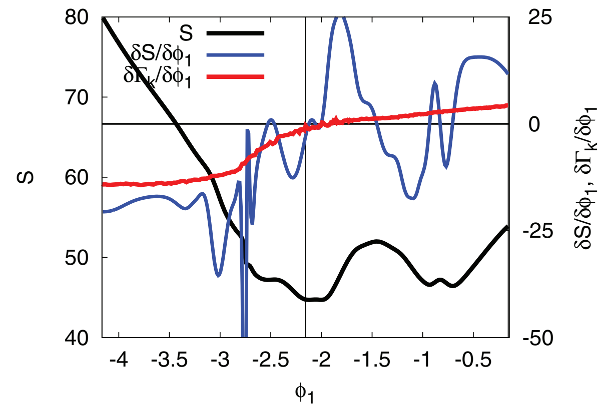

FIG. 6. Action $S[\phi]$ (black curve), gradient of action $\delta S[\phi]/\delta\phi_1$ (blue curve), and effective gradient of action $\delta\Gamma_k[\phi]/\delta\phi_1$ (red curve) for the Lorenz model. The true value for the data assimilation problem is $\phi_1 = -2.156$ and the first guess is -0.156.

D. Slow and fast degrees of freedom

Toward the application to high-dimensional systems, we briefly note a possible procedure for treating two distinct spatial modes in the coarse-grained data assimilation. Assume the infrared filter has a sharp cutoff in the momentum representation

$$R_k(p) = \begin{cases} a^2 & \text{if } |p| < k \\ 0 & \text{otherwise} \end{cases} \tag{81}$$

and the fluctuation χ can be decomposed into

$$\chi(p) = \begin{cases} \chi_s(p) & \text{if } |p| < k \\ \chi_f(p) & \text{otherwise.} \end{cases} \tag{82}$$

Then, as the mass $a^2 \to \infty$, we have

$$e^{-\Delta S_k[\chi_s + \chi_f]} = e^{-(a^2/2)\chi_s^T\chi_s} \to \delta[\chi_s], \tag{83}$$

which leads to

$$\frac{\delta\Gamma_k}{\delta\phi}[\phi] = \frac{\int d\chi_s \int d\chi_f \frac{\delta S}{\delta\phi}[\phi + \chi_s + \chi_f] \exp\left(-S[\phi + \chi_s + \chi_f] + S[\phi] + \frac{\delta\Gamma_k}{\delta\phi}[\phi](\chi_s + \chi_f)\right)\delta[\chi_s]}{\int d\chi_s \int d\chi_f \exp\left(-S[\phi + \chi_s + \chi_f] + S[\phi] + \frac{\delta\Gamma_k}{\delta\phi}[\phi](\chi_s + \chi_f)\right)\delta[\chi_s]} \tag{84}$$

$$\to \frac{\int d\chi_f \frac{\delta S}{\delta\phi}[\phi + \chi_f] e^{-S[\phi+\chi_f]+S[\phi]+(\delta\Gamma_k[\phi]/\delta\phi)\chi_f}}{\int d\chi_f e^{-S[\phi+\chi_f]+S[\phi]+(\delta\Gamma_k[\phi]/\delta\phi)\chi_f}}. \tag{85}$$

Hence, the sensitivity can be calculated as the average under the weight that integrates out the fast degrees of freedom, which will contribute to reducing the dimensionality of the path space. In the case of coupled atmosphere-ocean systems, we can assume that the atmospheric system is represented by χ_f and the oceanic system is χ_s. This suggests that, through this coarse-graining procedure, the sensitivity regarding the coupled system can be expressed by the slow oceanic variables alone.

E. Two-scale Lorenz model

To illustrate the application to multiscale systems, as mentioned in Sec. V D, we examine the sensitivity appearing during data assimilation in the two-scale Lorenz model [16]. This system can be written for the slow ζ and the fast ξ variables as

$$\frac{d\zeta_{l_1}}{dt} = \zeta_{l_1-1}(\zeta_{l_1+1} - \zeta_{l_1-2}) - \zeta_{l_1} + F - \frac{hc}{b}\sum_{l_2=1}^{M_2}\xi_{l_2,l_1}, \quad l_1 = 1,2,\ldots,M_1 \tag{86}$$

$$\frac{d\xi_{l_2,l_1}}{dt} = cb\xi_{l_2+1,l_1}(\xi_{l_2-1,l_1} - \xi_{l_2+2,l_1}) - c\xi_{l_2,l_1} + \frac{hc}{b}\zeta_{l_1}, \quad l_2 = 1,2,\ldots,M_2,\ l_1 = 1,2,\ldots,M_1 \tag{87}$$

$$\zeta_0 = \zeta_{M_1}, \quad \zeta_{M_1+1} = \zeta_1, \quad \zeta_{-1} = \zeta_{M_1-1}, \tag{88}$$

$$\xi_{0,l_1} = \xi_{M_2,l_1-1}, \quad \xi_{M_2+1,l_1} = \xi_{1,l_1+1}, \quad \xi_{M_2+2,l_1} = \xi_{2,l_1+1}, \quad l_1 = 1,2,\ldots,M_1, \tag{89}$$

where $\zeta(t) \in \mathbb{R}^{M_1}$ and $\xi(t) \in \mathbb{R}^{M_1M_2}$. We write the time evolution operator of the state $\theta = (\zeta^T,\xi^T)^T$ from the initial condition as

$$\Psi^{(j)}[\theta(t=0)] = \theta(t=t_j), \tag{90}$$

with time step $t_{j+1} - t_j = \Delta t$. If we choose the initial condition $\theta(t=0) = \phi \in \mathbb{R}^M$, where $M = M_1 + M_1M_2$, as the control variable, we can define a cost function as follows:

$$S[\phi] = \frac{1}{2\sigma_0^2}\|\phi - m_0\|^2 + \sum_{j=0}^{J-1}\frac{1}{2\gamma^2}\|y_{j+1} - \Psi^{(j+1)}[\phi]\|^2, \tag{91}$$

where $\|\cdot\|$ is the Euclidean norm. The parameters are set to $F = 6$, $J = 110$, $M_1 = 5$, $M_2 = 3$, $h = 1.6$, $b = c = 10$, $\sigma_0 = 2$, and $\gamma = 2$. With these parameters, this model is in an unstable regime with the first Lyapunov exponent $\lambda_1 \simeq 9.0 > 0$. The true initial condition $v_0 \in \mathbb{R}^M$ is given by model integration within an interval from a randomly chosen initial condition. The observation is sampled from a model sequence starting from v_0, with observational noise added to the sample. The observation is defined only at the times $j+1 = 3,6,\ldots,108$ and for the slow variables $l_1 = 1,2,\ldots,M_1$. The first guess m_0 is given by changing only the first component of v_0:

$$(m_0)_1 = (v_0)_1 + \sigma_0, \quad (m_0)_2 = (v_0)_2, \ldots, \quad (m_0)_M = (v_0)_M. \tag{92}$$

Similar to Eq. (81), we set the filter R_k as

$$R_k = \begin{cases} k_\zeta^2 & \text{for } \zeta(t=0) \\ k_\xi^2 & \text{for } \xi(t=0), \end{cases} \tag{93}$$

with $k_\zeta = 0.01^{-1}$ and $k_\xi = 0.04^{-1}$. This setting $k_\zeta > k_\xi$ mainly integrates out the fast degrees of freedom $\xi(t=0)$, whose typical fluctuation is assumed to be $\mathrm{var}(\xi(t=0)) \simeq k_\xi^{-2} = 0.04^2$. The time evolution of (86) and (87) as well as their adjoints are solved by the Runge-Kutta method with time step $\Delta t = 0.006$. As in Sec. V C, the experiments are designed to investigate the changes in the action, the gradient of action, and the effective gradient of action if we move the control variable as

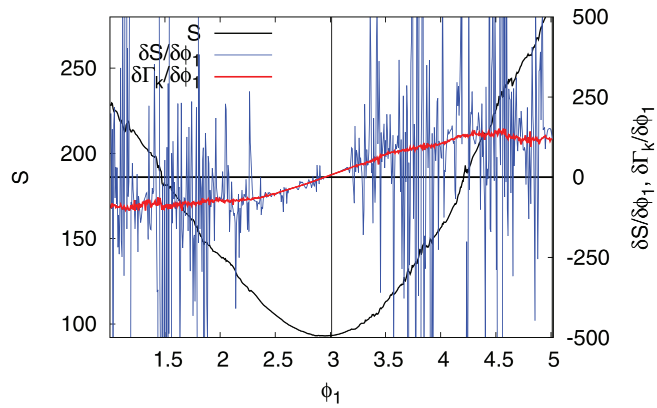

FIG. 7. Action $S[\phi]$ (black curve), gradient of action $\delta S[\phi]/\delta\phi_1$ (blue curve), and effective gradient of action $\delta\Gamma_k[\phi]/\delta\phi_1$ (red curve) for the two-scale Lorenz model. The true value for the data assimilation problem is $\phi_1 = 3.011$ and the first guess is 5.011.

$$\phi_1 = (v_0)_1 + \sigma_0\eta, \quad \phi_2 = (v_0)_2, \ldots, \quad \phi_M = (v_0)_M, \tag{94}$$

where $-1 \leqslant \eta \leqslant 1$.

Figure 7 shows the action (black curve), the gradient of action (blue curve), and the effective gradient of action (red curve) for this experiment. The action is almost parabolic in shape with many small bumps due to the fast degrees of freedom. Consequently, the gradient has many zeros that can complicate the variational data assimilation using it. However, the effective gradient has only one zero near the true value, similar to the result obtained for the Logistic map and for the one-scale Lorenz model. Thus, this example shows the potential application of the effective gradient for finding the optimal initial condition in multiscale data assimilation.

VI. CONCLUSION

We have investigated the use of the effective average action and its gradient in multiscale data assimilation. A framework has been proposed that allows the multiscale data assimilation problem to be solved by replacing the cost function with the effective average action. We have also proposed a method of evaluating the gradient of the effective average action numerically. In principle, this enables a broader range of data assimilation problems to be solved by seeking the stationary point of this effective average action numerically using its gradient.

This work can be summarized as follows:

(i) The concept of the effective average action provides a consistent framework for data assimilation in nonlinear multiscale systems.

(ii) If we can numerically evaluate the path integral

$$\frac{\delta\Gamma_k}{\delta\phi}[\phi] = \frac{\int d\chi \frac{\delta S}{\delta\phi}[\phi+\chi]e^{-\mathfrak{R}[\phi,\chi]}}{\int d\chi e^{-\mathfrak{R}[\phi,\chi]}} \equiv \left\langle \frac{\delta S}{\delta\phi}[\phi+\chi]\right\rangle_{\mathfrak{R}},$$

$$\mathfrak{R}[\phi,\chi] \equiv S[\phi+\chi] - S[\phi] - \frac{\delta\Gamma_k}{\delta\phi}[\phi]\chi + \Delta S_k[\chi],$$

with reasonable accuracy and computational burden, then we obtain the coarse-grained sensitivity, which constitutes a key factor in the variational data assimilation of a coarse-grained field.

(iii) The proposed procedure estimates the coarse-grained sensitivity through the Metropolis method by averaging an ensemble of original sensitivities that are distributed according to weights related to the nonlinearity:

$$\left\langle \frac{\delta S}{\delta \phi}[\phi + \chi] \right\rangle_{\mathfrak{R}} \simeq \frac{1}{N} \sum_{n=1}^{N} \frac{\delta S}{\delta \phi}[\phi + \chi^{(n)}].$$

(iv) The stationary problem for the effective average action $\Gamma_k[\phi]$ can be solved using a gradient method with the gradient $\delta\Gamma_k[\phi]/\delta\phi$.

(v) The stationary value $\hat{\phi}$ represents the extremum for the coarse-grained field after integrating out the ultraviolet fluctuations. This can be regarded as a solution of the multiscale data assimilation problem.

(vi) We demonstrated the usefulness of the effective gradient for data assimilation in a simple setting with the double-well potential, the logistic map, the one-scale Lorenz model, and the two-scale Lorenz model.

Future research should consider the following issues:

(a) The infrared filter works well when we can separate slow and fast modes cleanly, as described in Secs. V D and V E. However, in general, we have to deal with control variables that have continuous spectra. For such cases, we should carefully consider an infrared filter design that is suitable for revealing the slow dynamics of the system under consideration.

(b) In the case of a larger system, the computational burden of the Metropolis method could be huge, because we require many samples to yield a statistically reasonable integration result. Moreover, this should be incorporated into a recursive procedure or a fixed point calculation.

Despite these technical difficulties, the coarse-grained sensitivities are of great importance since they provide an invaluable perspective on the slow dynamics of multiscale systems. It should be noted that our approach in the present form has a fairly limited scope of application to data assimilation problems in geoscience, which typically require higher-dimensional systems.

ACKNOWLEDGMENTS

The author is grateful to H. D. I. Abarbanel for motivating this research and for several suggestions. This work is supported by Japan Agency for Marine-Earth Science and Technology (JAMSTEC). The numerical simulations were performed on the JAMSTEC SC supercomputer system.

APPENDIX A: PROPERTIES OF Γ_k

1. Perturbation expansion

To evaluate the path integral (50) through a perturbation expansion, we apply the approximation $\delta\Gamma_k[\phi]/\delta\phi \simeq \delta S[\phi]/\delta\phi$ in the exponent of Eq. (49) and truncate the Taylor series expansion of $S[\phi + \chi]$ to the quadratic order:

$$\begin{aligned} &S[\phi+\chi] - S[\phi] - \frac{\delta \Gamma_k}{\delta\phi}[\phi]\chi + \Delta S_k[\chi] \\ &\simeq S[\phi] + \frac{\delta S}{\delta\phi}[\phi]\chi + \frac{1}{2}\chi^T \frac{\delta^2 S}{\delta\phi^2}[\phi]\chi \\ &- S[\phi] - \frac{\delta S}{\delta\phi}[\phi]\chi + \frac{1}{2}\chi^T R_k \chi \\ &= \frac{1}{2}\chi^T \frac{\delta^2 S}{\delta\phi^2}[\phi]\chi + \frac{1}{2}\chi^T R_k\chi = \frac{1}{2}\chi^T\left(\frac{\delta^2 S}{\delta\phi^2}[\phi] + R_k\right)\chi. \end{aligned} \tag{A1}$$

Taking the Gaussian integral, we obtain

$$\begin{aligned} \Gamma_k[\phi] &\simeq S[\phi] - \ln \int d\chi \exp\left[-\frac{1}{2}\chi^T\left(\frac{\delta^2 S}{\delta\phi^2}[\phi] + R_k\right)\chi\right] \\ &= S[\phi] + \frac{1}{2}\ln\det\left(\frac{\delta^2 S}{\delta\phi^2}[\phi] + R_k\right). \end{aligned} \tag{A2}$$

Thus, we require at least the second derivative of the action $S[\phi]$ for the perturbation calculation of $\Gamma_k[\phi]$.

The same procedure can be applied to the gradient (53) as follows:

$$\begin{aligned} \frac{\delta\Gamma_k}{\delta\phi}[\phi] &\simeq \frac{\delta S}{\delta\phi}[\phi] + \frac{1}{2}\left\langle \chi^T\left(\frac{\delta^3 S}{\delta\phi^3}[\phi]\right)\chi\right\rangle_{\mathfrak{R}} \\ &= \frac{\delta S}{\delta\phi}[\phi] + \frac{1}{2}\mathrm{tr}\left\{\frac{\delta^3 S}{\delta\phi^3}[\phi]\left(\frac{\delta^2 S}{\delta\phi^2}[\phi] + R_k\right)^{-1}\right\}. \end{aligned} \tag{A3}$$

Hence, we require at least the second and third derivatives of the action $S[\phi]$ for the perturbation calculation of the gradient of $\Gamma_k[\phi]$.

2. Relationship to the growth of instabilities

We consider the case where $S[\phi]$ is the cost function of strong-constraint 4D-Var [17]. The meaning of $\ln\det$ in Eq. (A2) can be clarified by considering the basis of singular vectors. With the singular values $\sigma_1 > \sigma_2 > \cdots$, we can write

$$\frac{\delta^2 S}{\delta\phi^2}[\phi] \simeq \left(\frac{\delta F}{\delta\phi}[\phi]\right)^T\left(\frac{\delta F}{\delta\phi}[\phi]\right) = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots). \tag{A4}$$

Using a positive constant $k \gg 1$, we can define the infrared filter as

$$R_k = \begin{cases} k^2 & \text{if } \sigma_i < k \\ 0 & \text{otherwise.} \end{cases} \tag{A5}$$

Then we have

$$\frac{\delta^2 S}{\delta\phi^2}[\phi] + R_k \simeq \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, k^2, \ldots, k^2), \tag{A6}$$

$$\frac{1}{2}\ln\det\left(\frac{\delta^2 S}{\delta\phi^2}[\phi] + R_k\right) \simeq \sum_{\sigma_i \geqslant k}\ln\sigma_i + \sum_{\sigma_i < k}\ln k. \tag{A7}$$

That is, the term $\ln\det$ represents the sum of the logarithms of the leading singular values. The additional term in Eq. (A7) has the effect of integrating out the growing disturbances in the cost function.

APPENDIX B: ALGORITHM FOR COARSE-GRAINED DATA ASSIMILATION

Algorithm 1 coarse-grained_data_assimilation

```
Ψ_0 ← v_0                                                          ▷ set true value
for t = 0 → T − 1 do
    Ψ_{t+1} ← ψ(Ψ_t)
    generate ξ ∼ N(0, I)
    y_{t+1} ← Ψ_{t+1} + γξ                                         ▷ set observation
end for
φ_0 ← m_0                                                          ▷ set first guess
for i = 0 → I − 1 do                                               ▷ assimilation loop
    CALL_COST(y_{1:T}, φ_i, Ψ_{0:T}, S_i)
    CALC_SENSITIVITY(y_{1:T}, Ψ_{0:T}, ∇S_i)                       ▷ see gradient_calculations
    CALC_COARSE-GRAINED_SENSITIVITY(y_{1:T}, φ_i, S_i, ∇S_i, ∇Γ_i) ▷ see gradient_calculations
    if |∇Γ_i| < C_th then
        return φ_i, S_i
    end if
    φ_{i+1} ← φ_i − α∇Γ_i                                          ▷ update control variable
end for

procedure CALC_COST(y_{1:T}, φ_i, Ψ_{0:T}, S_i)
    Ψ_0 ← φ_i                                                      ▷ background term
    S_i ← 1/(2σ_0^2) |Ψ_0 − m_0|^2
    for t = 0 → T − 1 do
        Ψ_{t+1} ← ψ(Ψ_t)                                           ▷ forward time stepping
        S_i ← S_i + 1/(2γ^2) |y_{t+1} − Ψ_{t+1}|^2                 ▷ observational term
    end for
end procedure
```

Algorithm 2 gradient_calculations

```
procedure CALC_SENSITIVITY(y_{1:T}, Ψ_{0:T}, ∇S_i)
  Ψ̂_{0:T} ← 0
  for t = T − 1 → 0 do
    Ψ̂_{t+1} ← Ψ̂_{t+1} − (1/γ²)(y_{t+1} − Ψ_{t+1})        ▷ observational term
    Ψ̂_t ← Ψ̂_t + (∂ψ/∂Ψ_t)^T Ψ̂_{t+1}                      ▷ adjoint time stepping
    Ψ̂_{t+1} ← 0
  end for
  Ψ̂_0 ← Ψ̂_0 + (1/σ_0²)(Ψ_0 − m_0)                        ▷ background term
  ∇S_i ← Ψ̂_0
  Ψ̂_0 ← 0
end procedure

Procedure CALC_COARSE-GRAINED_SENSITIVITY(y_{1:T}, φ_i, S_i, ∇S_i, ∇Γ_i)
  ∇Γ_i ← ∇S_i                                            ▷ first guess of gradient in weight
  for l = 1 → L do                                       ▷ successive correction of gradient in weight
    χ ← 0
    ∇S ← ∇S_i
    R ← 0
    ∇R ← 0
    ∇Γ^acc ← 0
    for n = 0 → N − 1 do                                 ▷ Markov-chain loop
      generate ξ ~ 𝒩(0, I)
      χ* ← χ − ½σ²∇R + σξ                                ▷ proposed
      CALC_COST(y_{1:T}, φ_i + χ*, Ψ_{0:T}, S*)
      CALC_SENSITIVITY(y_{1:T}, Ψ_{0:T}, ∇S*)
      R* ← S* − S_i − ⟨∇Γ_i, χ*⟩ + ½k²|χ*|²
      ∇R* ← ∇S* − ∇Γ_i + k²χ*
      q_+ ← (1/2σ²)|χ − ½σ²∇R − χ*|²
      q_− ← (1/2σ²)|χ* − ½σ²∇R* − χ|²
      a ← −R* + R − q_− + q_+
      generate ζ ~ 𝒰(0,1)
      if a ⩾ 0 or ζ < exp(a) then                        ▷ Metropolis criterion
        χ ← χ*
        ∇S ← ∇S*
        R ← R*
        ∇R ← ∇R*
      end if
      ∇Γ^acc ← ∇Γ^acc + ∇S                               ▷ accumulate gradient
    end for
    ∇Γ_i ← ∇Γ^acc/N                                      ▷ take average
  end for
end procedure
```


[1] M. Bocquet, L. Wu, and F. Chevallier, Q. J. R. Meteor. Soc. **137**, 1340 (2011).

[2] M. Peña and E. Kalnay, Nonlin. Process. Geophys. **11**, 319 (2004).

[3] H. D. Abarbanel, Phys. Lett. A **373**, 4044 (2009).

[4] H. D. Abarbanel, *Predicting the Future: Completing Models of Observed Complex Systems* (Springer, Berlin, 2013).

[5] The convexity of $W[J]$ follows from Hölder's inequality $E[\exp((1-\gamma)J_1^T\chi+\gamma J_2^T\chi)]\leqslant E[\exp(J_1^T\chi)]^{1-\gamma}E[\exp(J_2^T\chi)]^{\gamma}$, $0\leqslant\gamma\leqslant 1$, where $E[G[\chi]]\equiv\int d\chi\, G[\chi]\exp(-S[\chi])/\int d\chi \exp(-S[\chi])$.

[6] A. Zee, *Quantum Field Theory in a Nutshell*, 2nd ed. (Princeton University Press, Princeton, 2010), Chap. VI.

[7] C. Wetterich, Phys. Lett. B **301**, 90 (1993).

[8] We use the following matrix-form notation to denote the spatiotemporal (or momentum space) integration of fields:

$$\int d^d x \sum_n J_n(x)\chi_n(x) \to J^T \chi,$$

$$\frac{1}{2}\int \frac{d^d q}{(2\pi)^d} \sum_{m,n} \chi_m(q)(R_k)_{m,n}(q)\chi_n(-q) \to \frac{1}{2}\chi^T R_k \chi,$$

$$\frac{\delta S}{\delta\phi(x)}[\phi] \to \frac{\delta S}{\delta\phi}[\phi],$$

where d is the dimension of the space-time and m and n are the indices of the components.

[9] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[10] W. K. Hastings, Biometrika **57**, 97 (1970).

[11] G. O. Roberts and R. L. Tweedie, Bernoulli **2**, 341 (1996).

[12] In particular, strong-constraint 4D-Var has

$$F[\phi] = \begin{bmatrix} B^{-1/2}(\phi - m_0) \\ R^{-1/2}(y - \Psi[\phi]) \end{bmatrix},$$

where Ψ is the model, B and R are the background and observational error covariance matrices, respectively, and m_0 and y are the first-guess field and observational data, respectively.

[13] R. M. May *et al.*, Nature (London) **261**, 459 (1976).

[14] K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation: A Mathematical Introduction* (Springer, Berlin, 2015).

[15] In this case, $F[\phi] = [(\phi - m_0)/\sigma_0, (y_1 - \Psi^{(1)}[\phi])/\gamma, \ldots, (y_J - \Psi^{(J)}[\phi])/\gamma]^T$.

[16] E. Lorenz, *Proceedings of the Annual Seminar on Predictability* (ECMWF, Shinfield Park, 1995), Vol. 1, pp. 1–18.

[17] A. C. Lorenc and T. Payne, Q. J. R. Meteor. Soc. **133**, 607 (2007).