Nonlinearity **17** (2004) 1867–1889 PII: S0951-7715(04)75330-X

Ruelle's linear response formula, ensemble adjoint schemes and Lévy flights

G L Eyink[1], **T W N Haine**[2] **and D J Lea**[3]

[1] Department of Applied Mathematics and Statistics, 3400 North Charles Street, The Johns Hopkins University, Baltimore, MA 21218-2682, USA
[2] Department of Earth and Planetary Sciences, 3400 North Charles Street, The Johns Hopkins University, Baltimore, MA 21218-2682, USA
[3] Applied Physics Laboratory, The Johns Hopkins University, 11100 Johns Hopkins Road, Laurel, MD 20723-6099, USA

E-mail: eyink@ams.jhu.edu

Received 23 January 2004, in final form 3 June 2004
Published 12 July 2004
Online at stacks.iop.org/Non/17/1867
doi:10.1088/0951-7715/17/5/016

Recommended by V Baladi

Abstract

A traditional subject in statistical physics is the linear response of a molecular dynamical system to changes in an external forcing agency, e.g. the Ohmic response of an electrical conductor to an applied electric field. For molecular systems the linear response matrices, such as the electrical conductivity, can be represented by *Green–Kubo formulae* as improper time-integrals of 2-time correlation functions in the system. Recently, Ruelle has extended the Green–Kubo formalism to describe the statistical, steady-state response of a 'sufficiently chaotic' nonlinear dynamical system to changes in its parameters. This formalism potentially has a number of important applications. For instance, in studies of global warming one wants to calculate the response of climate-mean temperature to a change in the atmospheric concentration of greenhouse gases. In general, a *climate sensitivity* is defined as the linear response of a long-time average to changes in external forces. We show that Ruelle's linear response formula can be computed by an *ensemble adjoint* technique and that this algorithm is equivalent to a more standard ensemble adjoint method proposed by Lea, Allen and Haine to calculate climate sensitivities.

In a numerical implementation for the 3-variable, chaotic Lorenz model it is shown that the two methods perform very similarly. However, because of a power-law tail in the histogram of adjoint gradients their sum over ensemble members becomes a *Lévy flight*, and the central limit theorem breaks down. The law of large numbers still holds and the ensemble-average converges to the desired sensitivity, but only very slowly, as the number of samples is increased. We discuss the implications of this example more generally for ensemble adjoint techniques and for the important practical issue of calculating climate sensitivities.

PACS numbers: 05.45.Ac, 92.70.Gt, 05.70.Ln, 05.40.Fb

(Some figures in this article are in colour only in the electronic version)

1. Introduction

A standard problem in the statistical mechanics of molecular systems is the study of their *linear response to an external field* [1, 2]. For example, a metallic conductor subjected to an applied electric field $\boldsymbol{E}$ will develop a current $\boldsymbol{J}$. In general, this current is given to good accuracy by Ohm's Law, $\boldsymbol{J} = \boldsymbol{\sigma}\boldsymbol{E}$, where $\boldsymbol{\sigma}$ is the so-called electrical conductivity. A key theoretical result is the *Green–Kubo formulae* for the conductivity and other similar linear response matrices [3, 4]. These formulae give the conductivity as a time-integral of a 2-time correlation function:

$$\boldsymbol{\sigma} = \int_0^\infty \mathrm{d}t \langle \boldsymbol{j}(t)\boldsymbol{j}^\top(0)\rangle_{(0)}, \tag{1}$$

where $\boldsymbol{j}(t)$ is the microscopic (fluctuating) electric current variable at time t and the average $\langle\cdot\rangle_{(0)}$ is over the Gibbsian thermal equilibrium ensemble at zero applied field. Such formulae are a standard tool to evaluate transport coefficients by molecular dynamics simulations [2].

Most early work has considered such a near-equilibrium response for Hamiltonian molecular dynamics. It is only recently that the linear response formulae have been generalized to any 'sufficiently chaotic', dissipative nonlinear dynamics with statistics far from thermodynamic equilibrium, in the work of Ruelle [5–8]. The problem he considered may be stated as follows: let the state of the dynamical system be denoted by a vector $\boldsymbol{x}$ and let f be a function of that state, an 'observable'. Also, let the dynamics be given by a differential equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{v}_{\boldsymbol{\alpha}}(\boldsymbol{x}(t)), \tag{2}$$

smoothly depending upon a set of parameters $\boldsymbol{\alpha}$. Finally, if $\boldsymbol{x}_{\boldsymbol{\alpha}}(t)$ is the solution of (2), let

$$\bar{f}^{\tau}_{\boldsymbol{\alpha}} = \frac{1}{\tau}\int_0^\tau \mathrm{d}t f(\boldsymbol{x}_{\boldsymbol{\alpha}}(t)), \tag{3}$$

denote the time-average of f. As long as the duration τ of the average is sufficiently long, then, by assumption of ergodicity, it coincides with an average over a steady-state ensemble. The latter will be denoted as $\langle f\rangle^*_{\boldsymbol{\alpha}}$. Then, the linear response is the change $\delta\langle f\rangle^*_{\boldsymbol{\alpha}}$ for a given change in the parameters $\delta\boldsymbol{\alpha}$ away from some fixed value $\boldsymbol{\alpha}_{(0)}$. For a sufficiently small change, $\delta\boldsymbol{\alpha} = \boldsymbol{\alpha} - \boldsymbol{\alpha}_{(0)}$, Ruelle established, under some assumptions, a linear relationship:

$$\delta\langle f\rangle^*_{\boldsymbol{\alpha}} \approx \boldsymbol{\sigma}^*_f \cdot \delta\boldsymbol{\alpha}, \tag{4}$$

where $\boldsymbol{\sigma}^*_f$ is a numerical matrix that depends upon $\boldsymbol{\alpha}_{(0)}$ but not upon $\delta\boldsymbol{\alpha}$. Ruelle showed that there is a formula like (1) for this response matrix, involving an improper integral of a 2-time correlation function.

An important problem in geophysics where Ruelle's formalism should apply is the calculation of the response of long-time climate averages to the change in some external parameter. For example, in studies of anthropogenic warming, one would like to calculate the response of the climate-mean global temperature to a change in the atmospheric concentrations of CO_2 or of other greenhouse gases. This is an example of a *climate sensitivity*. Clearly, calculating such a sensitivity falls exactly into the class of problems considered above and Green–Kubo formulae—in principle—should be valid. Of course, it is one thing to say that such formulae apply, and another matter to find a practical algorithm to apply them. As we show here, calculation of the general linear response formulae can be formulated as an *adjoint algorithm* [9]. Furthermore, there are close connections to an ensemble adjoint approach for climate sensitivity analysis proposed previously by Lea *et al* [10, 11]. Such algorithms have an important advantage over direct finite-difference methods, that they can yield the sensitivities of one variable to changes in several different control parameters simultaneously [9]. One of the purposes of this paper is to elaborate an ensemble adjoint algorithm to calculate climate sensitivity from the Ruelle formula and to compare it with the earlier, more standard one. In addition, the present results provide a firmer theoretical basis for the algorithm in [10, 11].

We find, in a numerical implementation for the 3-variable chaotic model of Lorenz [12], that both methods are very slowly converging in the number of samples, however, requiring an ensemble of immoderate size for even modest accuracy in the sensitivity. The difficulty is traced to certain rare dynamical trajectories in the Lorenz model which pass very close to the unstable fixed point at the origin in phase space, for which the adjoint gradient is enormous. These trajectories produce a *Lévy-type* power-law tail in the histogram of the adjoint gradients. As a consequence, the errors in the ensemble-average gradient are not governed by the standard central limit theorem, and are much more slowly decreasing than would be naïvely expected. It is unclear whether such difficulties will occur in applications of the ensemble adjoint algorithms to more realistic systems, but clearly the Lorenz model is a cautionary example. On this basis, we discuss the convergence properties of the algorithms more generally and the important practical issue of predictability of climate response.

The plan of this paper is as follows: in section 2 we describe in detail the general linear response formula of Ruelle. We discuss its dynamical foundations and regime of validity, based, in particular, upon a simple generalization to nonlinear response. In section 3, we introduce the two ensemble adjoint methods. We give no derivations there, but simply state the steps of the two algorithms. In section 4, we apply both methods numerically to the Lorenz model and compare results. In section 5, we discuss the convergence properties of the methods in general, under the assumption that adjoint gradients have a Lévy-type distribution. We also draw our final conclusions there. The two appendices contain some technical material. In appendix A we derive the generalization of Ruelle's formula to nonlinear response, and in appendix B we give a unified derivation of the Ruelle linear response formula and both ensemble adjoint algorithms.

2. The general linear response formula

As in the problem stated above, we consider an arbitrary perturbation of the general dynamical system (2) that is linear in the parameters around $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{(0)}$:

$$\delta \boldsymbol{v}(\boldsymbol{x}) := \left.\frac{\partial \boldsymbol{v}_\alpha}{\partial \boldsymbol{\alpha}}(\boldsymbol{x})\right|_{\alpha=\alpha_{(0)}} \cdot \delta\boldsymbol{\alpha}. \tag{5}$$

The Ruelle formula calculates the climate sensitivity of a state function $f(\boldsymbol{x})$ to such a change in the dynamics. The precise quantity is slightly different from that considered in the introduction, however. Let $P(\boldsymbol{x})$ be a probability distribution over the state space, representing an ensemble of initial data $\boldsymbol{x}$ of the system. Then let $P^t_\alpha(\boldsymbol{x}')$ be the distribution evolved forward under the dynamics, i.e. the distribution of the solutions $\boldsymbol{x}' = \boldsymbol{x}_\alpha(t; \boldsymbol{x})$ of the dynamics for initial data $\boldsymbol{x}$

distributed by $P(\boldsymbol{x})$. The Ruelle formula calculates the response in the average

$$\langle f\rangle_\alpha^t = \int \mathrm{d}\boldsymbol{x}' f(\boldsymbol{x}') P_\alpha^t(\boldsymbol{x}') = \int \mathrm{d}\boldsymbol{x} f(\boldsymbol{x}_\alpha(t;\boldsymbol{x})) P(\boldsymbol{x}).$$

In general, for 'reasonable' initial measures evolving under a 'suitably chaotic' dynamical system, the average, $\langle f\rangle_\alpha^t$, above is expected to converge as $t \to \infty$ to the same limit as the long-time average. In fact, there is expected to be a natural invariant measure $P_\alpha^*(\boldsymbol{x}')$ to which $P_\alpha^t(\boldsymbol{x}')$ tends for $t \to \infty$, in the sense that averages with respect to the latter converge to averages with respect to the former. This long-time stationary distribution represents the 'climate state' of the model. If the variables for which convergence of averages is valid include all bounded, continuous functions on the state space, then this corresponds to convergence to the invariant measure in the so-called *weak* sense. A condition of 'reasonableness' under which convergence should hold is that the initial measure $P(\boldsymbol{x})$ have a density with respect to the Liouville (Lebesgue) measure $\mathrm{d}\boldsymbol{x}$.[4] Furthermore, the time-average

$$\bar{f}_\alpha^\tau = \frac{1}{\tau}\int_0^\tau \mathrm{d}t f(\boldsymbol{x}_\alpha(t;\boldsymbol{x}))$$

will converge as $\tau \to \infty$ to the same limit, $\langle f\rangle_\alpha^* = \int \mathrm{d}\boldsymbol{x}'\ f(\boldsymbol{x}') P_\alpha^*(\boldsymbol{x}')$, for almost every initial point $\boldsymbol{x}$ chosen with respect to Liouville measure. For precise conditions on the dynamics under which such results are proved to hold, see [13], section IV.B. The essential requirement is one of uniform hyperbolicity of the dynamics, as for Anosov or axiom A systems. However, it is reasonable to conjecture that the same results will hold as well for a wide class of high-dimensional, chaotic dynamical systems, which are not uniformly hyperbolic but which are sufficiently complex that they behave 'as if' they were. This conjecture has been raised to the level of a general principle, the *chaotic hypothesis*, of Gallavotti and Cohen (see [14, 15] or [16, 17] for reviews).

Ruelle's general linear response formula [5, 8] gives $\delta\langle f\rangle^t$ for any small change $\delta\boldsymbol{v}$:

$$\delta\langle f\rangle^t = \int_0^t \mathrm{d}s \langle \boldsymbol{\kappa}_f^\top(s)\delta\boldsymbol{v}\rangle_{(0)}^{t-s}. \tag{6}$$

Here $\langle\cdot\rangle_{(0)}^{t-s}$ denotes that the ensemble-average over the initial data $\langle\cdot\rangle$ is evolved under the unperturbed dynamics ($\boldsymbol{\alpha} = \boldsymbol{\alpha}_{(0)}$) for a time $t - s$. The variable $\boldsymbol{\kappa}_f(t;\boldsymbol{x})$ is the response of $f(t;\boldsymbol{x}) := f(\boldsymbol{x}_{(0)}(t;\boldsymbol{x}))$ to a change of the initial datum $\boldsymbol{x}$, i.e.

$$\boldsymbol{\kappa}_f(t;\boldsymbol{x}) = \boldsymbol{\nabla}_{\boldsymbol{x}} f(t;\boldsymbol{x}). \tag{7}$$

We emphasize that the time-dependence of all expressions in this formula is calculated with zeroth-order dynamics. A linear response law follows, $\delta\langle f\rangle^t = \boldsymbol{\sigma}_f^t\cdot\delta\boldsymbol{\alpha}$, with

$$\boldsymbol{\sigma}_f^t = \int_0^t \mathrm{d}s \langle \boldsymbol{\kappa}_f^\top(s)(\boldsymbol{\nabla}_\alpha\boldsymbol{v})_{(0)}\rangle_{(0)}^{t-s}. \tag{8}$$

[4] To avoid excessively cumbersome notation that may be opaque to less mathematically-inclined readers, we have denoted a general measure on the state space by $P(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$, which seems to assume the existence of a density function $P(\boldsymbol{x})$ with respect to Lebesgue measure $\mathrm{d}\boldsymbol{x}$. Of course, this will not generally be true, e.g. the natural invariant measure $P^*(\boldsymbol{x})$ for a dissipative dynamics is itself generally supported on a strange attractor and is singular with respect to Lebesgue measure. In any such case, our notation should be interpreted as a convenient replacement for a more mathematically correct $P(\mathrm{d}\boldsymbol{x})$, which assumes existence of no density. We trust that our notational shorthand will cause no confusion.

The formula (6) simplifies considerably if we take the distribution over the initial data to be $P^*_{(0)}(\boldsymbol{x})$, the stationary distribution of the unperturbed dynamics. In that case[5],

$$\delta\langle f\rangle^t = \int_0^t \mathrm{d}s\langle \boldsymbol{\kappa}_f^\top(s)\delta\boldsymbol{v}\rangle^*_{(0)}. \tag{9}$$

As there is no longer a dependence upon $t-s$, we can also express the result using a cumulative response function $\boldsymbol{K}_f(t;\boldsymbol{x}) = \int_0^t \mathrm{d}s\, \boldsymbol{\kappa}_f(s;\boldsymbol{x})$. Then

$$\delta\langle f\rangle^t = \langle \boldsymbol{K}_f^\top(t)\delta\boldsymbol{v}\rangle^*_{(0)}.$$

As described above, Ruelle's formula for the 'finite-time' linear response is rigorously true for any dynamical system with smooth dependence upon a parameter $\boldsymbol{\alpha}$. No assumption of chaoticity or even ergodicity is required. However, in order to describe the response of the invariant measure $P^*_{\boldsymbol{\alpha}}$, the limit $t\to\infty$ must be considered and then some such hypotheses are needed. Formally, one expects that the formula (9) will hold approximately at large t for any 'nice' choice of the initial distribution $P(\boldsymbol{x})$ when the dynamics is 'sufficiently chaotic', because then in (6) $P^{t-s}_{(0)}(\boldsymbol{x}) \approx P^*_{(0)}(\boldsymbol{x})$ (in the weak sense) as $t\to\infty$. Therefore, one can plausibly expect that the responses of averages in the invariant measure are given by

$$\delta\langle f\rangle^* = \int_0^\infty \mathrm{d}t\langle \boldsymbol{\kappa}_f^\top(t)\delta\boldsymbol{v}\rangle^*_{(0)}. \tag{10}$$

This formula has been rigorously proved by Ruelle for differentiation of SRB measures of uniformly hyperbolic systems [5]. It is an exact analogue to the Green–Kubo formula (1) for electrical conductivity. Indeed, the classical Green–Kubo formula for conductivity has been derived by means of (10) for a deterministic dynamical model of a conductor in [18, 19]. This work predated and helped to motivate Ruelle's general theory.

In line with the chaotic hypothesis [14, 15] one might expect the Ruelle formula (10) for the 'infinite-time' response to hold for a wide class of chaotic dynamical systems with many degrees of freedom, even those which are not uniformly hyperbolic. To strengthen this expectation, we present here an 'axiomatic' derivation of Ruelle's formula which should make clearer exactly which dynamical properties are sufficient for its validity. It is proved in proposition 1 of appendix A that Ruelle's formula (10) can be generalized to a *nonlinear response*, as

$$\Delta\langle f\rangle^* = \int_0^\infty \mathrm{d}t\langle \boldsymbol{\kappa}_{f_{\boldsymbol{\alpha}}}^\top(t)\cdot\Delta\boldsymbol{v}\rangle^*_{(0)}, \tag{11}$$

where $\Delta\langle f\rangle^* = \langle f\rangle^*_{\boldsymbol{\alpha}} - \langle f\rangle^*_{(0)}$ and $\Delta\boldsymbol{v}(\boldsymbol{x}) = \boldsymbol{v}_{\boldsymbol{\alpha}}(\boldsymbol{x}) - \boldsymbol{v}_{(0)}(\boldsymbol{x})$ are finite variations, not infinitesimal ones. For other notations, see appendix A. The necessary and sufficient condition for validity of (11) is that $P^*_{(0)}$ should lie in the (weak convergence) domain of attraction of $P^*_{\boldsymbol{\alpha}}$ for initial measures evolved under the dynamics with parameter $\boldsymbol{\alpha}$. Now suppose that the integrands in this nonlinear response formula are *integrable uniformly in* $\boldsymbol{\alpha}$, in the sense that

$$\left|\left\langle \boldsymbol{\kappa}_{f_{\boldsymbol{\alpha}}}^\top(t)\left(\frac{\Delta\boldsymbol{v}}{\Delta\boldsymbol{\alpha}}\right)\right\rangle^*_{(0)}\right| \leqslant I(t) \tag{12}$$

with $\Delta\boldsymbol{\alpha} = \boldsymbol{\alpha} - \boldsymbol{\alpha}_{(0)}$, for some function satisfying $\int_0^\infty \mathrm{d}t\, I(t) < \infty$, for all $\boldsymbol{\alpha}$ in some open neighbourhood of $\boldsymbol{\alpha}_{(0)}$. Then Ruelle's formula (10) follows by a straightforward

[5] Our notation differs slightly from that of Ruelle [8]. He defines instead a linear map $\kappa(t)$ from vector fields $\delta\boldsymbol{v}(\boldsymbol{x})$ into linear functionals on the space of observables f, related to our $\boldsymbol{\kappa}_f(t)$ by

$$[\kappa(t)\delta\boldsymbol{v}](f) = \langle \boldsymbol{\kappa}_f^\top(t)\delta\boldsymbol{v}\rangle^*_{(0)}.$$

Thus, Ruelle's response function includes ours for all possible choices of f, recognizing that the dependence of $\boldsymbol{\kappa}_f(t)$ upon f is a simple linear one. Note that Ruelle also considered in [8] a somewhat more general situation than the one we study here, in which the perturbation to the dynamics $\delta\boldsymbol{v}(\boldsymbol{x}, t)$ could also be time-dependent.

application of Lebesgue's dominated convergence theorem. We may state this result as follows:

Theorem 1. *Consider invariant measures P^*_α of a dynamics (2) that depend smoothly on a parameter α. Ruelle's differentiation formula (10) holds whenever (i) $P^*_{(0)}$ lies in the (weak) domain of attraction of P^*_α under the flow $x_\alpha(t)$ and (ii) the integrability condition (12) holds, both conditions valid for all α in an open neighbourhood of $\alpha_{(0)}$.*

The proof of this theorem, based on the nonlinear response formula, is the same as that by which Ohm's Law was originally established for the Lorenz model of conduction [18, 19], and was used even earlier to discuss the validity of transport laws in stochastic lattice gases [20]. As this result should make clear, conditions much weaker than uniform hyperbolicity are sufficient for the validity of Ruelle's formula (10). The chaotic hypothesis is used only to justify the two conditions (i), (ii) of theorem 1. Such properties as the decay of correlations in (12) are still hard to prove rigorously for realistic dynamical systems, without uniform hyperbolicity. However, they are certainly more generally valid. To be sure, (i) and (ii) must fail in some cases, e.g. at certain bifurcations where differentiability of the invariant measures really fails. Nevertheless, our theorem 1 shows that formula (10) gives the linear response under conditions that are close to necessary for it to make sense. An idea of this sort seems to be what Ruelle had in mind when he wrote, '[if the] response for physical systems has a simple expression, it must be the one given by our formal calculations' [8].

In a practical numerical application of Ruelle's formula (10), one cannot (of course!) take the limit $t \to \infty$ to obtain the 'climate sensitivity'. However, one need not actually take the limit to infinity. Any time t greater than that, $\tau^*_{(0)}$, for which the integrand $\langle \kappa_f^\top(t)\delta v\rangle^*_{(0)}$ drops sensibly to zero will suffice to give an accurate result. As long as there is rapid decay of correlations, as we assume, then such a finite time exists, for any desired degree of approximation. Intuitively, the period $\tau^*_{(0)}$ represents a relaxation time to the steady-state or a spin-up time to climate statistics.

3. Two ensemble adjoint methods

An ensemble adjoint algorithm to calculate climate sensitivity was first proposed and studied by Lea *et al* [10, 11]. The chief advantage of an adjoint approach —in contrast to more direct methods—is that it allows one to calculate the sensitivity to a variety of different dynamical perturbations δv with just a single integration of the system dynamics and its adjoint [9]. This remains the main motivation to develop such a scheme.

3.1. The standard adjoint formulation

The approach of [10, 11] was based upon a standard adjoint method [9] to calculate the response of time-integrals, such as the following:

$$\bar{f}^\tau(x) = \frac{1}{\tau}\int_0^\tau \mathrm{d}t\, f(x(t; x)), \tag{13}$$

but then further averaged over an ensemble of initial data x. For this reason, we refer to the method of Lea *et al* as the 'standard adjoint (SA) method'. To begin their algorithm, one must have an N-sample ensemble of state vectors, x_n, $n = 1, \ldots, N$ which are distributed according to a chosen density $P(x)$. The precise steps of the Lea–Allen–Haine algorithm are as follows.

(i) For each $n = 1, \ldots, N$, solve $\dot{x} = v_{(0)}(x)$ with initial condition $x(0) = x_n$ and denote the solution as $x_n(t)$ for $0 \leqslant t \leqslant \tau$.

(ii) For each $n = 1, \ldots, N$, solve backward in time the adjoint system

$$\dot{\boldsymbol{\lambda}}(t) + \left(\frac{\partial \boldsymbol{v}_{(0)}}{\partial \boldsymbol{x}}\right)^{\top} (\boldsymbol{x}_n(t))\boldsymbol{\lambda}(t) = -(\boldsymbol{\nabla} f)(\boldsymbol{x}_n(t)) \tag{14}$$

with final condition $\boldsymbol{\lambda}(\tau) = \mathbf{0}$, and denote the solution as $\boldsymbol{\lambda}_n(t)$ for $0 \leqslant t \leqslant \tau$.

(iii) Estimate the climate sensitivity by

$$\delta\langle \bar{f}^{\tau}\rangle_N \equiv \frac{1}{N}\sum_{n=1}^{N} \frac{1}{\tau}\int_0^{\tau} \mathrm{d}t\, \boldsymbol{\lambda}_n^{\top}(t)\delta\boldsymbol{v}(\boldsymbol{x}_n(t)). \tag{15}$$

This consists simply of N independent repetitions of the standard adjoint algorithm, which gives

$$\overline{\delta f}^{\,\tau}(\boldsymbol{x}_n) = \frac{1}{\tau}\int_0^{\tau} \mathrm{d}t\, \boldsymbol{\lambda}_n^{\top}(t)\delta\boldsymbol{v}(\boldsymbol{x}_n(t)),$$

the time-averaged response with initial datum $\boldsymbol{x}_n$, for each $n = 1, \ldots, N$. These results are then averaged over the N-sample ensemble of initial data.

We emphasize the importance of this additional ensemble-averaging. As discussed in the previous section, it can be proved under some general assumptions that the time-average $\bar{f}^{\tau}(\boldsymbol{x})$ in (13) will converge as $\tau \to \infty$ to the ensemble-average $\langle f\rangle^*$ for almost every choice of initial data $\boldsymbol{x}$ with respect to Lebesgue measure ([13], section IV.B). However, it does not follow that $\delta\bar{f}^{\tau}(\boldsymbol{x})$ will converge to the desired climate sensitivity $\delta\langle f\rangle^*$. Indeed, it was shown by Lea *et al* [10, 11] in a numerical study of the 3-variable chaotic Lorenz model that $\delta\bar{f}^{\tau}(\boldsymbol{x})$ does not converge at all as $\tau \to \infty$, for particular perturbations $\delta\boldsymbol{v}$, state variables f, and general initial data $\boldsymbol{x}$ (see also our section 4). Rather, this quantity diverges to infinity at an exponential rate governed by the dominant Lyapunov exponent. It was found in [10, 11] that averaging over an N-sample ensemble, as above, delayed that divergence to later and later times τ, the greater the number of samples N employed. Using an intermediate time τ of the order of the spin-up time and an ensemble of $N = 299$ samples produced an estimate of the climate sensitivity accurate to about 10%, whereas a single adjoint calculation for a time-average of duration 299τ produced an estimate too large by 100 orders of magnitude!

3.2. The linear-response adjoint formulation

Interestingly, Ruelle's general linear response formula discussed in section 2 can also be cast as an ensemble adjoint algorithm to calculate climate sensitivity, which we call the linear-response adjoint (LRA) method. We discuss two versions of this algorithm: an instantaneous method and a method with additional time-averaging.

3.2.1. Instantaneous LRA algorithm. In the instantaneous method, one does the following:

(i′) For each $n = 1, \ldots, N$, solve $\dot{\boldsymbol{x}} = \boldsymbol{v}_{(0)}(\boldsymbol{x})$ with initial condition $\boldsymbol{x}(0) = \boldsymbol{x}_n$ and denote the solution as $\boldsymbol{x}_n(t)$ for $0 \leqslant t \leqslant \tau$.

(ii′) For each $n = 1, \ldots, N$ and each $0 \leqslant t \leqslant \tau$, solve $\dot{\boldsymbol{\kappa}}(s,t) + (\partial\boldsymbol{v}_{(0)}/\partial\boldsymbol{x})^{\top}(\boldsymbol{x}_n(s))\boldsymbol{\kappa}(s,t) = 0$ backward in time s with final condition $\boldsymbol{\kappa}(t,t) = (\boldsymbol{\nabla} f)(\boldsymbol{x}_n(t))$, and denote the solution as $\boldsymbol{\kappa}_n(s,t)$ for $0 \leqslant s \leqslant t \leqslant \tau$.

(iii′) Setting $\boldsymbol{\kappa}_n(t) = \boldsymbol{\kappa}_n(0, t)$, estimate the climate sensitivity by

$$\delta\langle f\rangle_N^\tau \equiv \int_0^\tau \mathrm{d}t \frac{1}{N}\sum_{n=1}^N \boldsymbol{\kappa}_n^\top(t)\delta\boldsymbol{v}(\boldsymbol{x}_n) = \frac{1}{N}\sum_{n=1}^N \boldsymbol{K}_n^\top(\tau)\delta\boldsymbol{v}(\boldsymbol{x}_n). \tag{16}$$

The final result is an approximation to the Green–Kubo formula (9). Thus, it must be assumed here that $P(\boldsymbol{x}) = P_{(0)}^*(\boldsymbol{x})$, i.e. the initial ensemble members $\boldsymbol{x}_n$, $n = 1, \ldots, N$ are chosen from the stationary distribution for the zeroth-order dynamics. Initial conditions of this sort may be obtained, for example, by running the zeroth-order dynamics with an arbitrary initial state in the domain of the stable attractor. The N samples may then be selected from this single trajectory $\boldsymbol{x}(t)$ as $\boldsymbol{x}_n = \boldsymbol{x}(n\tau_{(0)}^*)$, $n = 1, \ldots, N$, where $\tau_{(0)}^*$ is a typical relaxation time of the system. The whole set of time-histories $\boldsymbol{x}_n(t)$ for each $n = 1, \ldots, N$ and $0 \leqslant t \leqslant \tau$ must be stored for use in integrating the adjoint equations for $\boldsymbol{\kappa}_n(s, t)$ backward in s. However, the results of that integration need not be stored for $0 \leqslant s \leqslant t$, but instead only the output values $\boldsymbol{\kappa}_n(t) = \boldsymbol{\kappa}_n(0, t)$ are required. In fact, only the cumulative response $\boldsymbol{K}_n(t)$ needs to be stored. A detailed justification of the steps in this algorithm is given in appendix B.

3.2.2. Averaged LRA algorithm. The quantity calculated by the adjoint algorithm in the previous subsection is not the same as that calculated in the standard method. Instead, the LSA method yields the *instantaneous* or *fixed-time sensitivity*, as it appears in the Green–Kubo formula (9), with the average over $P_{(0)}^*(\boldsymbol{x})$ approximated by the N-sample average.

However, the LRA algorithm may be simply modified to give the time- and ensemble-averaged sensitivity, the same as the standard method. The steps (i″)–(iii″) of the averaged LRA algorithm are the same as (i′)–(iii′) above, except that an additional averaging is performed to give the *time-average cumulative response*:

$$\bar{\boldsymbol{K}}_n(\tau) \equiv \frac{1}{\tau}\int_0^\tau \mathrm{d}t\, \boldsymbol{K}_n(t).$$

The final result

$$\delta\langle \bar{f}^\tau\rangle_N = \frac{1}{N}\sum_{n=1}^N \bar{\boldsymbol{K}}_n^\top(\tau)\delta\boldsymbol{v}(\boldsymbol{x}_n). \tag{17}$$

is the same as in the standard algorithm (SA). In a numerical implementation of this algorithm it is convenient to save running totals of both $\boldsymbol{K}_n(t)$ and $t\bar{\boldsymbol{K}}_n(t) = \int_0^t \mathrm{d}s\, \boldsymbol{K}_n(s)$ at each intermediate time $0 \leqslant t \leqslant \tau$. However, as the final output of the algorithm, only $\bar{\boldsymbol{K}}_n(\tau)$ $n = 1, \ldots, N$ need be saved for a sufficiently large $\tau \approx \tau_{(0)}^*$.

4. Numerical results for the Lorenz model

Here, we illustrate and test the SA and LRA methods on a prototypical chaotic system. The Lorenz model [12] is given by

$$\begin{aligned} &\boldsymbol{x}(t) = [x(t), y(t), z(t)]^\top, \\ &\boldsymbol{v}_\alpha(\boldsymbol{x}) = [\sigma(y - x), rx - y - xz, -bz + xy]^\top, \\ &\boldsymbol{\alpha} = [\sigma, r, b]^\top. \end{aligned} \tag{18}$$

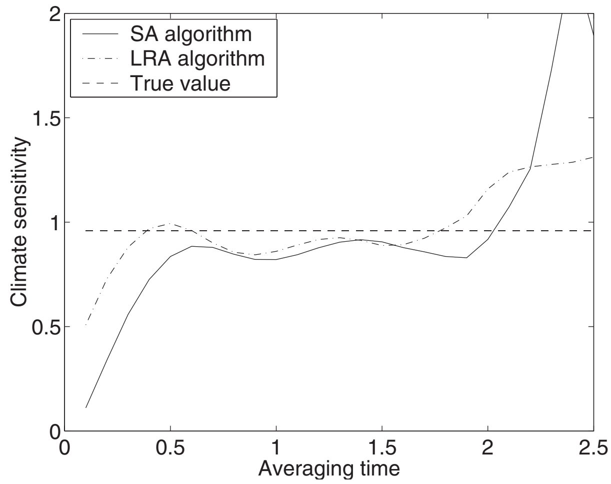

Figure 1. Climate sensitivities estimated using an ensemble adjoint algorithm plotted against averaging time, by the SA method (——) and the LRA method (— · —). The dashed line shows a direct estimate 0.96 of the climate sensitivity [10]. $N = 1.5 \times 10^8$ samples were used.

We consider sensitivity of z to perturbations in r. That is, $f = z$, $\delta\boldsymbol{\alpha} = [0, \delta r, 0]^\top$, and thus,

$$\frac{\partial \boldsymbol{v}_{(0)}}{\partial \boldsymbol{x}}(\boldsymbol{x}_{(0)}(t;\boldsymbol{x})) = \begin{bmatrix} -\sigma_{(0)} & \sigma_{(0)} & 0 \\ r_{(0)} - z_{(0)}(t) & -1 & -x_{(0)}(t) \\ y_{(0)}(t) & x_{(0)}(t) & -b_{(0)} \end{bmatrix},$$
$$(\boldsymbol{\nabla} f)(\boldsymbol{x}) = [0, 0, 1]^\top, \qquad (19)$$
$$\delta\boldsymbol{v}(\boldsymbol{x}) = [0, x\delta r, 0]^\top$$

for each ensemble member being considered. As in [10], the results of our adjoint calculations will be compared with those of a direct calculation, in which the average $\bar{z}$ over a long time-interval is evaluated for various values of r and then $\mathrm{d}\bar{z}/\mathrm{d}r|_{r_{(0)}}$ calculated by a finite-difference approximation. Important considerations on the proper choice of Δr for calculating the difference quotient were discussed in [10]. For any finite value of the averaging time τ, $\bar{z}^\tau$ appeared to have large derivatives with respect to r, whose magnitudes grew exponentially with τ. However, the difference quotients $\Delta\bar{z}^\tau/\Delta r$ had plateaux for a range of small values of Δr, which persisted to smaller Δr as τ was further increased. Therefore, it could be concluded on the basis of this numerical exercise that the limit $\tau \to \infty$ of $\bar{z}^\tau$ was indeed differentiable in r, with a finite derivative close to 0.96. For $r < 90$ the derivative appeared to fail to exist only near the value $r \approx 24.06$, which is the bifurcation point where a chaotic attractor appears in the Lorenz model (for details, see [10]). It is important to emphasize that the direct calculations performed there, as described, are the sole basis for our conviction that the derivative response $\mathrm{d}\langle z\rangle^*/\mathrm{d}r$ exists in the Lorenz model. Furthermore, they provide us with its numerical value, 0.9600 ± 0.0002, for comparison with the adjoint algorithms in this paper.

In figure 1 we show results from the SA algorithm for various values of τ. The forward equation (2) and backward equation (14) are integrated using a 4th-order Runge–Kutta scheme with a timestep $\delta t = 0.005$. In the case we consider, $N = 1.5 \times 10^8$, $\sigma_{(0)} = 10$, $b_{(0)} = 8/3$, and $r_{(0)} = 28$. For these parameters, the leading Lyapunov exponent $\lambda_1 \approx 0.906$, which gives a reasonable estimate of the spin-up timescale as $\tau^*_{(0)} \approx 1/\lambda_1 \approx 1.1$. For the times τ shown, the SA method gives an O(10–20%) accurate estimate of the sensitivity of $\bar{z}$ to changes in r

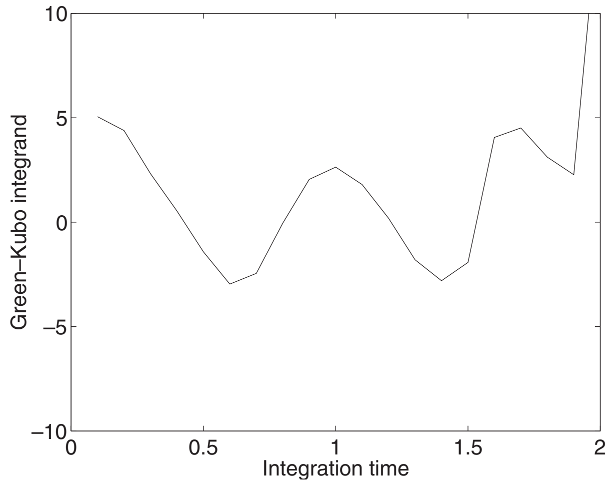

Figure 2. The Green–Kubo integrand in Ruelle's formula plotted against the integration time, estimated by an ensemble-average with $N = 1.5 \times 10^8$ samples.

(full line). Figure 1 also shows the corresponding results using the (averaged) LRA algorithm. We believe, incidentally, that this is the first numerical use of the Ruelle formula to calculate the linear response of a chaotic dynamical system. This method gives results virtually identical to those of the SA algorithm, as expected from our discussion in the appendix[6]. In particular, when $N = 1.5 \times 10^8$, the results of both methods begin to depart from the true sensitivity for $\tau \geqslant 2$, due to very large values in the solution of the adjoint equations for a few of the samples. The time of this divergence is delayed by using more ensemble members, or greater N. Further insight may be gained by plotting the integrand of the Ruelle formula as a function of time, as shown in figure 2. It is only for times $t \leqslant 1.5$ that $N = 1.5 \times 10^8$ is sufficiently large that the ensemble approximation to the integrand plotted there is convergent. As may be seen, the integrand is decaying, with oscillations of decreasing amplitude over the time interval [0, 1.5]. It is impossible to make a precise estimate of the rate of decay from the range of time available to us. Because of the agreement of the LRA sensitivity with the direct one, we expect that the decay is integrable in time, but even this cannot be confidently inferred from the data.

The above results are reassuring as tests of principle, but the number of ensemble members employed, $N = 1.5 \times 10^8$, is far too large to be practical in realistic applications. It is therefore

[6] They are not identical because of various approximations in our numerical implementation. The proof of the equivalence of SA and LRA algorithms in appendix B can be carried over also to the discrete-time dynamical systems obtained by numerical discretization of the Lorenz model ODEs. However, to obtain results equal to working precision, one must use the corresponding SA and LRA methods of the discrete systems. Instead, we independently discretized the equations of the continuous-time SA and LRA methods. Because we employed a high-order accurate time-discretization (4th-order Runge–Kutta), we did not need to use a consistent discretization of the two sets of equations, which would have guaranteed their exact equivalence. Furthermore, the proof of equivalence requires that the time-integral in the Green–Kubo formula (16) in the LRA method be discretized with the same time-increment δt as used in the Runge–Kutta integration of the ODEs. In fact, we employed a time-increment $\Delta t = 0.1$ in the Riemann sum approximation of the integral (16) which was considerably larger than the value $\delta t = 0.005$ which was used in the ODE solution. This is justified because the Green–Kubo integrand is smoother in time than individual realizations of the dynamics, and a coarser resolution is permitted to evaluate its time-integral. Without this additional approximation, the LRA method would be much more expensive than the SA method. A detailed comparison of the computational aspects of the two methods would be out of place here, but we just note that there is a trade-off in which the LRA method will generally be costlier than SA in terms of number of flops but will drastically reduce certain long-term memory or storage costs.

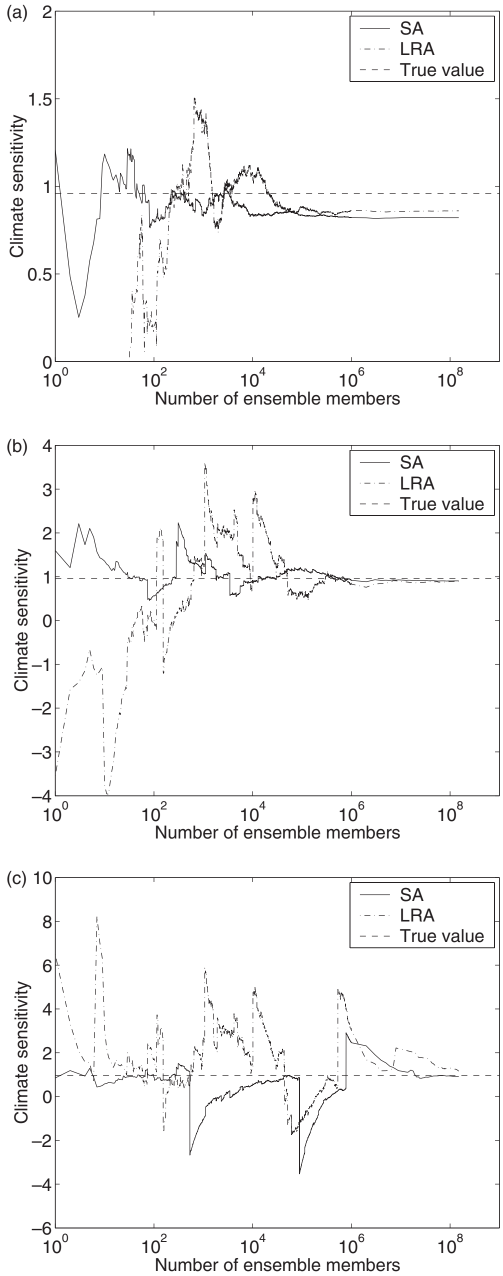

Figure 3. Climate sensitivities by the SA and LRA methods plotted against the number of samples N, for averaging times (*a*) $\tau = 1.0$, (*b*) $\tau = 1.5$ and (*c*) $\tau = 2.0$. Solid and dot-dashed lines are for SA and LRA, respectively. The dashed line shows the direct estimate 0.96 of climate sensitivity [10]. Note the change in ordinate.

important to consider the rate of convergence in N. In figures 3(*a*)–(*c*) we plot the sensitivities by both SA and LRA algorithms for three times $\tau = 1.0$, 1.5 and 2.0, as functions of N. The results are very similar for the two methods. Most strikingly, the graphs are often punctuated by very large jumps arising from the addition of a single new sample. The envelopes of the

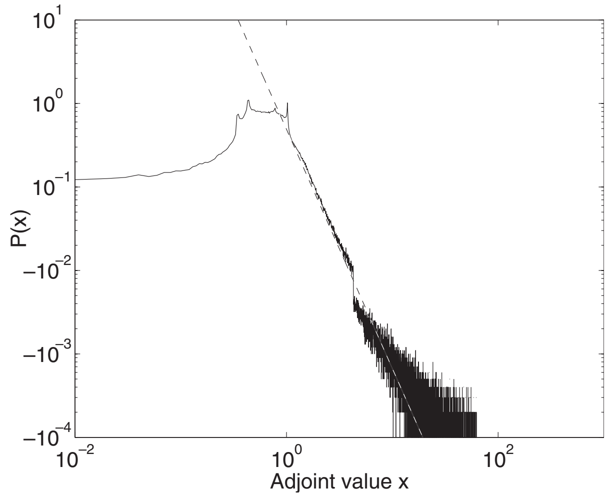

Figure 4. A histogram of the SA adjoint gradient values x for averaging time $\tau = 1.0$ from $N = 10^6$ samples, plotted against x (——). The dashed line is the best power-law fit to the tail $\sim Ax^{-(1+\alpha)}$ with $A = 0.4897$, $\alpha = 1.865$.

curves also decay surprisingly slowly. According to our best fit (see below), the decay is $\sim N^{\beta}$ with $\beta = -0.4637$, -0.4449 and -0.1562 for $\tau = 1.0$, 1.5 and 2.0, respectively. At all these times the decay exponent is smaller than for typical Monte Carlo, in which errors go as $\sim N^{-1/2}$ based on the central limit theorem. Furthermore, the decay appears to become much slower as τ is increased. As it is necessary to take $\tau \gg \tau^*_{(0)}$ to obtain the climate statistics, this implies that the sample size N necessary to obtain the long-time sensitivity will be quite large.

To get some insight into this slow decay, we plot in figure 4 the histogram of the individual terms in the sum over the N-sample SA ensemble for $\tau = 1.0$

$$\mathcal{S}_N = \sum_{n=1}^{N} \overline{\delta z}^{\tau}(\boldsymbol{x}_n). \tag{20}$$

We see that this distribution has a long power-law tail $(\overline{\delta z}^{\tau})^{-(1+\alpha)}$ with $\alpha \approx 1.865$. It follows immediately from this fact that the variance fails to be finite, as required for validity of the central limit theorem. The N-sample average, while expected to converge on the basis of the law of large numbers, will do so very slowly. The size of the errors may be estimated from the Gnedenko–Doeblin theorem [21], which generalizes the central limit theorem to random variables whose distributions exhibit such fat power-law tails for $0 < \alpha < 2$. The values of a sum like that in (20) perform a random walk with sporadic, long jumps, called a *Lévy flight*. Such random processes arise in fields as diverse as astronomy, physics, biology, economics and communication engineering [22–25]. It is typical of these processes that the maximum term in the N-sample sum, or the maximum step-size, scales as $N^{1/\alpha}$, the same as the magnitude of the entire sum. Rescaled as $(\mathcal{S}_N - \mu N)/N^{1/\alpha}$, with μ the mean, the sum converges as $N \to \infty$ to a random variable X_α distributed according to a *Lévy stable law*. In the theorem of Gnedenko–Doeblin the summands are taken to be statistically independent, but one would assume that the same result will hold so long as correlations decay sufficiently rapidly. For example, this behaviour has been observed in the output of chaotic maps of the real line whose invariant measures exhibit such fat tails [26]. It can then be expected that the ensemble mean

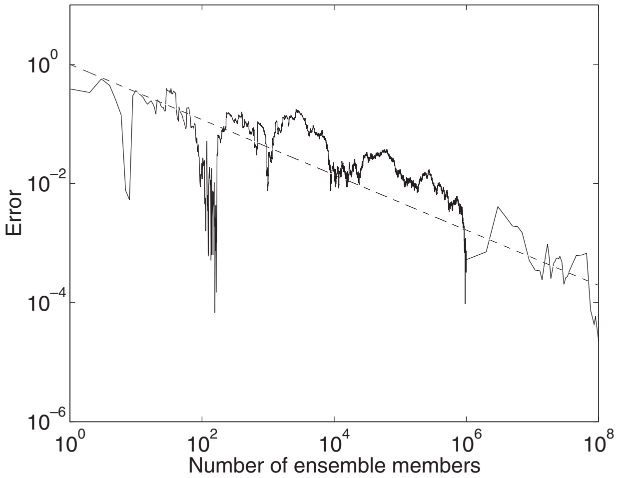

Figure 5. The error in the SA climate sensitivity from ensemble adjoints for averaging time $\tau = 1.0$, plotted against number of samples N (——). The dashed line is the fit by $(N/N_0)^{\beta}$ with $N_0 = 1$, $\beta = -0.4637$.

goes asymptotically as

$$\frac{1}{N}\mathcal{S}_N \sim \langle \bar{\delta z}^{\tau} \rangle + N^{(1/\alpha)-1} X_{\alpha}.$$

For $1 < \alpha < 2$ the law of large numbers still holds and the ensemble mean converges to the desired sensitivity, albeit slowly. The value $\alpha = 1.865$ obtained from the distribution in figure 4 nicely accounts for the slow decay of errors observed in figure 3(a). This is seen more clearly in figure 5 which plots the error $\varepsilon_N = |(1/N)\mathcal{S}_N - \langle \overline{\delta z}^{\tau} \rangle|$ versus the number of samples N in logarithmic coordinates. Also plotted there is a straight line with slope $\beta = 1/\alpha - 1 = -0.4637$, for α determined from the tail of the histogram in figure 4. It is clear that the decay of the error at large N is consistent with $\varepsilon_N \sim N^{\beta}$ for $\beta = 1/\alpha - 1$.

The results shown in figures 4 and 5 are for $\tau = 1.0$, but similar results hold at other times. Indeed, the tail of the histogram of δz seems to get fatter for increasing τ, with $\alpha = 1.802$ for $\tau = 1.5$ and $\alpha = 1.185$ for $\tau = 2.0$. Correspondingly, the rate of convergence of the N-sample ensemble-means in (15)–(17) becomes slower, the errors decaying approximately as N^{β}, with $\beta = (1/\alpha) - 1$ giving $\beta = -0.4449$ for $\tau = 1.5$ and $\beta = -0.1562$ for $\tau = 2.0$. For the latter time, $\tau = 2.0$, we plot the histogram of the adjoint gradients in figure 6 and the sample-mean error ε_N versus the number of samples N in figure 7. The tail of the histogram is so fat that the mean value just barely exists and the decay of the error in N is extremely slow, consistent with a fit by $\varepsilon_N \sim N^{\beta}$ for $\beta = 1/\alpha - 1 = -0.1562$. It is remarkable that the decay rate drops so precipitously in going from $\tau = 1.0 - 1.5$ to $\tau = 2.0$. For $\tau = 1.0 - 1.5$, the decay exponents β are only slightly smaller in magnitude than the value -0.5 implied by the central limit theorem. It is suggestive that this drastic change in the decay rate occurs at a time τ of the same order as the spin-up time $\tau^*_{(0)} = 1.1$, as estimated from the leading Lyapunov exponent. Integration of the adjoint linear equations backward in time should produce strong exponential divergences for times of that order.

In order to gain some insight into the dynamics which produces the fat tails, we have studied the size of the adjoint gradients produced by various initial data on the attractor. In figure 8 we plot the points in a 1 million member ensemble, colour coded according to the magnitude of the adjoint gradient for $\tau = 2.0$ that results from each initial condition (projected in a

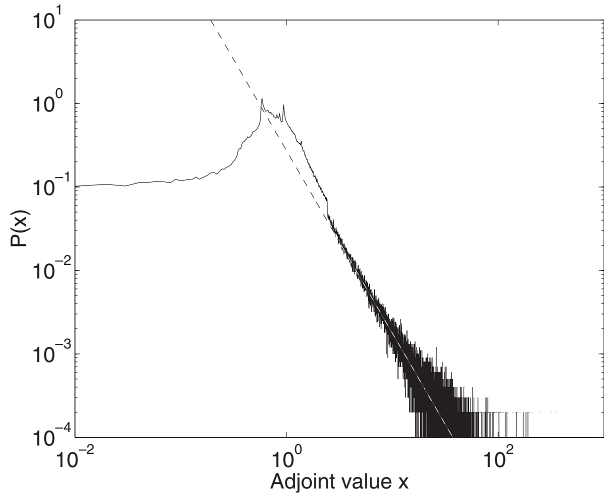

Figure 6. A histogram of the SA adjoint gradient values x for averaging time $\tau = 2.0$ from $N = 10^6$ samples, plotted against x (——). The dashed line is the best power-law fit to the tail $\sim Ax^{-(1+\alpha)}$ with $A = 0.2718$, $\alpha = 1.185$.

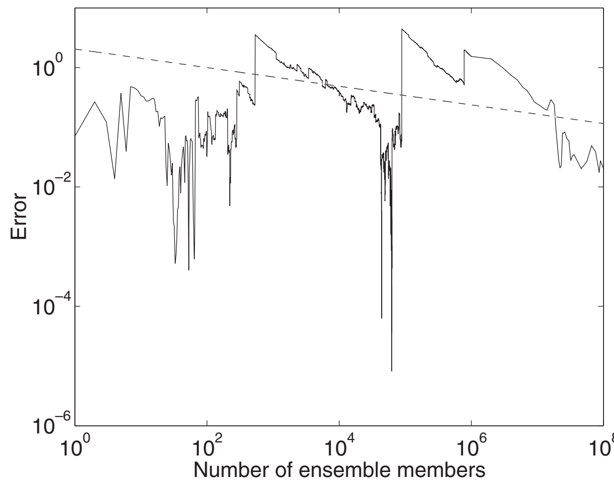

Figure 7. The error in the SA climate sensitivity, from ensemble adjoints for averaging time $\tau = 2.0$, plotted against number of samples N (——). The dashed line is the fit by $(N/N_0)^{\beta}$ with $N_0 = 10^2$, $\beta = -0.1562$.

two-dimensional x–z surface). The highest values are situated along clear tracks within the attractor. The largest adjoint, in this ensemble, is indicated by a cross. It arises from an initial condition

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0.543\,432\,224 \\ -0.352\,672\,725 \\ 20.004\,069\,000 \end{bmatrix} \tag{21}$$

and the value of the resulting adjoint gradient is $\mathrm{d}\bar{z}^{\tau}/\mathrm{d}r = 3.931\,214 \times 10^5$. The phase trajectory for this initial condition shows a behaviour which is very typical of all the initial

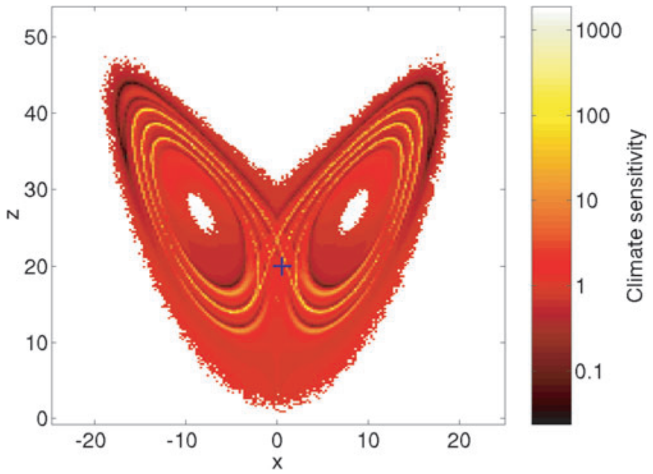

Figure 8. Plot of the average SA adjoint sensitivity as a function of x–z location (in 0.25×0.25 squares) for a 1 million member ensemble with $\tau = 2.0$. The cross shows the location of the highest adjoint value in the ensemble.

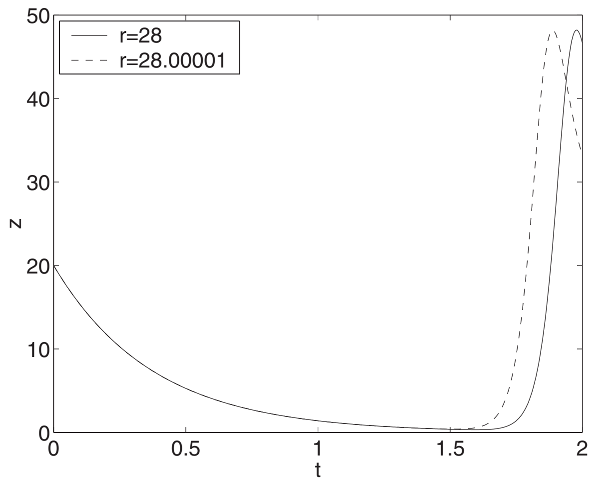

Figure 9. Trajectory $z(t)$ as a function of t for the highest-adjoint-value initial condition of the 1 million member ensemble, $(x, y, z) = (0.543\,432\,224, -0.352\,672\,725, 20.004\,069\,000)$. The solid line is for $r = 28$ and the dashed line shows the results for r perturbed by $\delta r = 10^{-5}$.

conditions which produce large adjoint gradients. All such phase orbits approach the unstable fixed point $\mathbf{0} = (0, 0, 0)$ of the Lorenz model, close to its two-dimensional stable manifold $W^s(\mathbf{0})$. In fact, the tracks in figure 8 seem to consist of points on the Lorenz attractor which lie sufficiently close to that stable manifold. A recent numerical computation of $W^s(\mathbf{0})$ shows that it contains three 'scrolls' that wind through the attractor at locations very close to the tracks in our figure 8 (see [27], figure 3). For a typical trajectory leading to a large gradient the three components, x, y and z, each decrease to very small values, until the flow along the unstable direction suddenly expels the phase point away from the origin. In figure 9 we plot the trajectory $z(t)$ as a function of t for the initial condition (21), where the solid line is for $r = 28$ and the dashed line is for r perturbed by $\delta r = 10^{-5}$. Comparing these two trajectories, we can use a finite-difference approximation to see the source of the large time-averaged z-gradient that results for this initial condition. Such trajectories that pass close to the origin are well known to be quite rare. For example, simple estimates ([28], appendix F) show that, of all trajectories passing through the plane $z = r - 1$, only about one in a million passes through a unit cube centred at the origin, for the standard choice of parameter values $r = 28$, $b = 8/3$, $\sigma = 10$. The probability that a trajectory will make an even closer approach to the origin rapidly becomes very small [28]. Rare rajectories of this type that closely approach the origin, where there is extreme sensitivity to parameters, are the cause of the fat tail in the distribution of adjoint gradients for the Lorenz model. As the existence of the Lorenz attractor and some of its properties have recently been rigorously established [29,30], it is possible that the results found here numerically may be accessible to a more analytical investigation.

5. Discussion and conclusions

The results that have been presented in the preceding section seem somewhat discouraging about the prospects for a practical adjoint algorithm to calculate climate sensitivities. At least for the Lorenz model, the convergence of such N-sample ensemble methods as $N \to \infty$ is very slow, due to a fat tail in the distribution of the adjoint gradients for time-averages. The convergence rate is essentially the same for the LRA method, based on Ruelle's formula, as for the standard adjoint (SA) algorithm. In each case, errors decay only as $\sim N^{\beta}$ with $-\frac{1}{2} < \beta < 0$, slower than for typical Monte Carlo. Assuming that the averaging time $\tau = 2.0$ is large enough to obtain the long-time climate sensitivity, we can use our fit to the error in figure 7, i.e. $(N/N_0)^{\beta}$ with $N_0 = 100$ and $\beta = -0.1562$, in order to estimate the size of the ensemble required to attain a given accuracy. From this formula, $N \approx 2.5 \times 10^8$ would be required to obtain the sensitivity to 10% accuracy, while an ensemble of size $N \approx 6.4 \times 10^{14}$ would be required for 1% accuracy! Clearly, such requirements make the algorithms totally impractical, at least for the Lorenz model. It is not clear that the fat tails which we have encountered there will also appear in more realistic geophysical models. Therefore, we still believe that the methods show some promise, but further study is required to see how widespread are such tail-phenomena. A good first step would be to consider some other standard chaotic dynamical systems to see whether similar difficulties will occur there. For example, the Rössler model [31] does not have its two dynamical fixed points imbedded in the strange attractor, whereas the Lorenz model attractor is the closure of the unstable manifold $W^u(\mathbf{0})$ of the fixed point at the origin. Therefore, we expect that the Rössler system may be quite different in its response behaviour. However, admittedly, on the basis of our experience in the Lorenz model alone, we would have little confidence in the results of an ensemble calculation with even several thousand samples. In many cases, e.g. for realistic climate general circulation models, ensembles of only a few hundred samples will be available.

In addition to the issue of convergence as $N \to \infty$, the convergence as $\tau \to \infty$ must also be considered. It is shown in appendix B of this article that the SA algorithm will give the same result as the LRA method based upon Ruelle's formula, if one takes limits in the order $N \to \infty$ first and then $\tau \to \infty$. Assuming for the moment that $N = \infty$, the rate of convergence in time τ should, therefore, be similar for the two methods. In the LRA algorithm, a value $\tau \approx \tau^*_{(0)}$ will generally suffice for the Green–Kubo time-integral to achieve its ultimate value. The size of the error will depend upon the rate of decay of the integrand. If this is a power law $\sim(t/\tau^*_{(0)})^{-(1+\gamma)}$ for some $\gamma > 0$, then the error will go as $\sim(\tau^*_{(0)}/\tau)^{\gamma}$ in τ. However, with N finite, it is a difficult problem to determine an optimal choice τ_N to give the most accurate estimate of the climate sensitivity. As follows from the results in [10, 11], taking τ as large as possible for finite N is not a good choice. Instead, an intermediate value τ_N was found to work better in [10, 11]. At the moment, we have no good method to estimate such a value τ_N *a priori*. The only statement we can make is that, if $N \to \infty$, then it should be true that the optimal $\tau_N \to \infty$ as well.

Although the results of our study are somewhat negative regarding practicality of the proposed ensemble adjoint algorithms, nevertheless they have thrown considerable light on theoretical issues. Our demonstration in the appendix of the equivalence of the SA algorithm to the LRA method, based on Ruelle's formula, gives a fundamental justification to the former method, which was proposed by Lea *et al* [10, 11] on a more heuristic and pragmatic basis. We can also clarify some important issues regarding the smoothness of the steady-state response of chaotic dynamical systems. For example, it was speculated in [32] that the steady-state average $\bar{z}$ in the Lorenz model might be a continuous but nowhere-differentiable function of the parameter r. If so, then there would be no linear response at all, for some range of the parameter r. In fact, it is known that even for molecular dynamical systems in thermal equilibrium, linear response may fail. For example, the famous long-time tails discovered in hard-sphere fluids by Alder and Wainwright [33] by molecular dynamics simulations imply a decay of the Green–Kubo integrands as $t^{-d/2}$ in d space-dimensions. Thus, ordinary linear response behaviour must break down in molecular systems for $d = 2$, as the Green–Kubo formulae for the transport coefficients diverge logarithmically in that case. For dimensions $d \geqslant 3$, the Green–Kubo integrals converge and linear response is valid. However, the higher order nonlinear response terms in a Taylor expansion in the perturbation involve multiple time-integrals [4], and these generally diverge even for $d \geqslant 3$. Therefore, the first-derivative response at equilibrium exists but the full response is non-analytic in the perturbation. As an example, mode-coupling theory for hard-sphere fluids [34] predicts that the Newtonian linear relation between shear stress e and strain γ fails in two-dimensions, with $e \propto |\gamma| \ln |\gamma|$, and that the higher order response is non-analytic in three-dimensions, with $e = -\eta\gamma + c|\gamma|^{3/2}$ (where η is the shear viscosity). This is a simple instance of what is sometimes called *Dorfman's lemma:* the dissipative, steady-state response of molecular systems is expected to be a non-analytic function of all control parameters [2].

Less is known about the steady-state response of general chaotic nonlinear dynamical systems. In addition to the linear response formula derived by Ruelle in [8], it is also possible to calculate formally the higher order terms in an *assumed* analytic expansion, yielding multiple time-integral formulae [7]. The general expectation is that, if the time-integrals in those expressions converge up to a given order, then the corresponding higher order response law is valid, with coefficients given by those formulae. The question, therefore, becomes one of the decay of time-correlations, such as that in the formula (10). If the decay is too slow to be integrable, then the response is non-differentiable. Assuming this slow decay holds for an open domain of the parameter $\boldsymbol{\alpha}_{(0)}$, then, over that entire domain, the response will be like a Weierstrass continuous, nowhere-differentiable function of $\boldsymbol{\alpha}_{(0)}$, as conjectured in [32]. However, if there is a good analogy with transport behaviour for molecular systems in three space dimensions and above, then the first-order derivative response is valid and non-analyticity appears only at higher orders. Our numerical result on the integrand of Ruelle's formula for the variable z in the Lorenz model, plotted in figure 2, can hardly resolve the issue in that case. However, the 10% agreement that we found between the results of both the SA and LRA methods and the direct calculation does seem to suggest that these methods converge as $N \to \infty$ and then $\tau \to \infty$. We, therefore, believe that the integrand plotted in figure 2 has a decay in time which is integrable. In that case, $\bar{z}$ in the Lorenz model, as a function of r, may behave like the integral of a Weierstrass function and not as a Weierstrass function itself.

Even if we are correct that the infinite-time average of z in the Lorenz model is differentiable in r and that its derivative is given by Ruelle's formula, the results of this work cast a doubt on the practical utility of this sort of differential mean response quantity for predicting the response of real-world climate to changes in dynamical parameters. Although the PDF peaks in our figures 4 and 6 are near the ensemble-average sensitivity there are long power-law tails to the right, with very large gradients. The realizations lying in those tails will show a response very much larger than the ensemble-mean. While the ensemble-mean in the Lorenz model appears to depend smoothly on the parameter r, we human beings do not have the luxury of averaging over even a small number of possible worlds. Instead, we are interested in the future of the one world in which we happen to live! Furthermore, we want to know the response of averages over some time period which might be long (centuries, millenia, etc) but not infinite. Ergodicity implies that such long-time averages ('climate') are statistically sharper and more predictable than instantaneous states ('weather'). It might therefore be naïvely assumed that *climate response* is also statistically sharp and does not vary much from realization to realization of the initial data. Unfortunately, the plots in our figures 4 and 6 show that the time-average response is far from predictable and becomes even less so as the period of averaging $\tau \to \infty$. Thus, there is a relatively high probability that an initial condition on the attractor has a response that is very different from the ensemble-average response given by Ruelle's formula. An important conclusion of our work is that the time-average climate response over long, but finite, time-intervals is inherently unpredictable in the Lorenz model. It remains to be seen how generally this is true.

Acknowledgments

GE acknowledges NSF grant DMS-0113649 for support of this work. TH was supported by NSF grant OCE-0136327. DL is a Johns Hopkins University Applied Physics Laboratory postdoctoral scholar.

Appendix A. Generalization of Ruelle's formula to nonlinear response

Let us start with the invariant measure $P^*_{(0)}$ of the unperturbed dynamics and consider its evolution P^t_α to time t under the dynamics with parameter $\boldsymbol{\alpha}$. We consider the difference

$$\Delta\langle f\rangle^t := \langle f\rangle^t_\alpha - \langle f\rangle^*_{(0)}.$$

We shall prove first the following formula which generalizes the finite-time linear response result (9) of Ruelle to finite-time nonlinear response:

$$\Delta\langle f\rangle^t = \int_0^t \mathrm{d}s \langle \boldsymbol{\kappa}^\top_{f_\alpha}(s)\cdot\Delta\boldsymbol{v}\rangle^*_{(0)}. \tag{22}$$

Here, $\Delta\boldsymbol{v}(\boldsymbol{x}) = \boldsymbol{v}_\alpha(\boldsymbol{x}) - \boldsymbol{v}_{(0)}(\boldsymbol{x})$. Also, for $f \in C^1$ we have set $f_\alpha(t;\boldsymbol{x}) = f(\boldsymbol{x}_\alpha(t;\boldsymbol{x}))$ and defined its response to a change of the initial datum $\boldsymbol{x}$ as

$$\boldsymbol{\kappa}_{f_\alpha}(t;\boldsymbol{x}) = \boldsymbol{\nabla}_x f_\alpha(t;\boldsymbol{x}). \tag{23}$$

The key step is the following lemma.

Lemma 1.

$$\frac{\mathrm{d}}{\mathrm{d}t} f_\alpha(t;\boldsymbol{x}) = \boldsymbol{\kappa}^\top_{f_\alpha}(t;\boldsymbol{x})\boldsymbol{v}_\alpha(\boldsymbol{x}).$$

Proof. By the group property of the flow under the dynamics, for $0 \leqslant s \leqslant t$,

$$\begin{aligned} f_\alpha(t;\boldsymbol{x}) &= f(\boldsymbol{x}_\alpha(t;\boldsymbol{x})) \\ &= f(\boldsymbol{x}_\alpha(s;\boldsymbol{x}_\alpha(t-s;\boldsymbol{x}))) \\ &= f_\alpha(s;\boldsymbol{x}_\alpha(t-s;\boldsymbol{x})). \end{aligned} \tag{24}$$

Taking the derivative with respect to t and using the chain rule gives

$$\frac{d}{\mathrm{d}t} f_\alpha(t;\boldsymbol{x}) = \boldsymbol{\kappa}_{f_\alpha}^\top(s;\boldsymbol{x}_\alpha(t-s;\boldsymbol{x}))\boldsymbol{v}_\alpha(\boldsymbol{x}_\alpha(t-s;\boldsymbol{x})).$$

The limit of this expression for $s \to t^-$ is the statement of the lemma. ∎

We now use lemma 1 and the fundamental theorem of calculus in order to write

$$f_\alpha(t;\boldsymbol{x}) - f(\boldsymbol{x}) = \int_0^t \mathrm{d}s\, \boldsymbol{\kappa}_{f_\alpha}^\top(s;\boldsymbol{x})\boldsymbol{v}_\alpha(\boldsymbol{x}).$$

Averaging over $\boldsymbol{x}$ with respect to $P_{(0)}^*$ gives

$$\Delta\langle f\rangle^t = \int_0^t \mathrm{d}s \langle \boldsymbol{\kappa}_{f_\alpha}^\top(s)\boldsymbol{v}_\alpha\rangle_{(0)}^*.$$

To complete the proof of (22), we need just observe that

$$\langle \boldsymbol{\kappa}_{f_\alpha}^\top(t)\boldsymbol{v}_{(0)}\rangle_{(0)}^* = 0 \tag{25}$$

for all $t > 0$. In fact, $\langle(\boldsymbol{v}_{(0)}\cdot\boldsymbol{\nabla})F\rangle_{(0)}^* = 0$ for all $F \in C^1$ is the differential condition of stationarity of $P_{(0)}^*$ under the unperturbed dynamics. Applying this condition for $F = f_\alpha(t)$ gives (25). Now, using the latter we see that

$$\langle \boldsymbol{\kappa}_{f_\alpha}^\top(t)\boldsymbol{v}_\alpha\rangle_{(0)}^* = \langle \boldsymbol{\kappa}_{f_\alpha}^\top(t)(\Delta\boldsymbol{v})\rangle_{(0)}^*$$

for all $t > 0$. Thus, the stated result (22) follows.

We now consider the limit $t \to \infty$ in the following proposition

Proposition 1. *Let* $\Delta\langle f\rangle^* := \langle f\rangle_\alpha^* - \langle f\rangle_{(0)}^*$. *Then, the weak limit* $\lim_{t\to\infty} P_\alpha^t = P_\alpha^*$ *if and only if for all bounded* $f \in C^1$

$$\Delta\langle f\rangle^* = \int_0^\infty \mathrm{d}t \langle \boldsymbol{\kappa}_{f_\alpha}^\top(t)\cdot\Delta\boldsymbol{v}\rangle_{(0)}^*.$$

The latter equality includes the statement that the improper integral converges.

Proof. The 'only if' part is a direct consequence of (22), since weak convergence $P_\alpha^t \to P_\alpha^*$ is equivalent to $\Delta\langle f\rangle^t \to \Delta\langle f\rangle^*$ for all bounded, continuous f. The 'if' part again uses (22) to show that $\Delta\langle f\rangle^t \to \Delta\langle f\rangle^*$ for all bounded $f \in C^1$. However, the latter are dense in C_b, the space of bounded, continuous functions on the state space with the uniform norm. ∎

Since $P_\alpha^t = P_{(0)}^* \circ \boldsymbol{x}_\alpha(-t)$, weak convergence $P_\alpha^t \to P_\alpha^*$ is a condition expressing 'evolutionary stability' of the invariant measures. The condition may be stated in terms of $\mathcal{D}(P_\alpha^*)$, the domain of attraction of P_α^* under weak convergence, or the set of probability measures which converge weakly under the dynamics $\boldsymbol{x}_\alpha(t)$ to P_α^* as $t \to \infty$. The condition is that $\mathcal{D}(P_\alpha^*) \ni P_{(0)}^*$, at least for $|\boldsymbol{\alpha} - \boldsymbol{\alpha}_{(0)}|$ not too large.

It might be expected that the equation (22) for the finite-time, nonlinear response can be evaluated numerically by an ensemble adjoint algorithm similar to that discussed in section 3.2 for the equation (9) giving the finite-time, linear response. As a matter of fact, such an adjoint algorithm may easily be constructed and it differs only slightly from the algorithm in section 3.2. The crucial difference is that in the new algorithm the adjoint equation solved backward in time in step (ii′) is for the perturbed dynamics with parameter $\boldsymbol{\alpha}$. Because of this, the primary advantage of an adjoint algorithm is lost, since the response functions in the modified algorithm must be recalculated for every new choice of the parameter $\boldsymbol{\alpha}$. Thus, there is no obvious advantage of an adjoint algorithm based on formula (22) compared with a direct calculation of the time-averages at two different values of the parameter $\boldsymbol{\alpha}$. In fact, from the results presented in the text it appears that formula (22) will be more costly to apply and will yield a less accurate result than a direct calculation.

Appendix B. Unified derivation of the two adjoint algorithms

Here, we present an elementary, unified derivation of the SA and LRA algorithms. Let $\boldsymbol{x}(t;\boldsymbol{x})$ be the solution of $\dot{\boldsymbol{x}}(t) = \boldsymbol{v}(\boldsymbol{x}(t))$ with initial condition $\boldsymbol{x}(0;\boldsymbol{x}) = \boldsymbol{x}$. Varying the equation for a perturbation of the dynamics $\boldsymbol{v}(\boldsymbol{x}) = \boldsymbol{v}_{(0)}(\boldsymbol{x}) + \delta\boldsymbol{v}(\boldsymbol{x})$ gives

$$\delta\dot{\boldsymbol{x}}(t;\boldsymbol{x}) = \boldsymbol{A}(t;\boldsymbol{x})\delta\boldsymbol{x}(t;\boldsymbol{x}) + \delta\boldsymbol{v}(\boldsymbol{x}_{(0)}(t;\boldsymbol{x})), \tag{26}$$

where

$$\boldsymbol{A}(t;\boldsymbol{x}) := \frac{\partial\boldsymbol{v}_{(0)}}{\partial\boldsymbol{x}}(\boldsymbol{x}_{(0)}(t;\boldsymbol{x})).$$

This has the formal integral solution

$$\delta\boldsymbol{x}(t;\boldsymbol{x}) = \int_0^t \mathrm{d}s\, \boldsymbol{G}_{(0)}(t,s;\boldsymbol{x})\delta\boldsymbol{v}(\boldsymbol{x}_{(0)}(s;\boldsymbol{x})), \tag{27}$$

where $\boldsymbol{G}_{(0)}(t,s;\boldsymbol{x})$ is the 2-time response function

$$\boldsymbol{G}_{(0)}(t,s;\boldsymbol{x}) = \mathrm{T}\exp\left(\int_s^t \mathrm{d}u\,\boldsymbol{A}(u;\boldsymbol{x})\right) \tag{28}$$

and T exp denotes the time-ordered exponential. Thus, the response function satisfies

$$[\partial_t - \boldsymbol{A}(t;\boldsymbol{x})]\boldsymbol{G}_{(0)}(t,s;\boldsymbol{x}) = 0 \tag{29}$$

for $t > s$ with initial condition $\boldsymbol{G}_{(0)}(s,s;\boldsymbol{x}) = \boldsymbol{I}$, the identity matrix. For later purposes we define also the anti-response (adjoint) function $\bar{\boldsymbol{G}}_{(0)}(s,t;\boldsymbol{x}) := [\boldsymbol{G}_{(0)}(t,s;\boldsymbol{x})]^\top$, which is represented by

$$\bar{\boldsymbol{G}}_{(0)}(s,t;\boldsymbol{x}) = \bar{\mathrm{T}}\exp\left(\int_s^t \mathrm{d}u\,\boldsymbol{A}^\top(u)\right), \tag{30}$$

where $\bar{\mathrm{T}}$ exp denotes the anti-time-ordered exponential. It satisfies the adjoint equation

$$[\partial_s + \boldsymbol{A}^\top(s)]\bar{\boldsymbol{G}}_{(0)}(s,t) = 0 \tag{31}$$

for $s \leqslant t$, with final condition $\bar{\boldsymbol{G}}_{(0)}(t,t;\boldsymbol{x}) = \boldsymbol{I}$.

Appendix B.1. Ruelle linear response formula

Appendix B.1.1. Fixed-time sensitivity. The response of $f(\boldsymbol{x}(t))$ to a perturbation of the dynamics is

$$\begin{aligned}\delta f(\boldsymbol{x}(t)) &= (\nabla f)^\top(\boldsymbol{x}_{(0)}(t))\delta\boldsymbol{x}(t)\\ &= \int_0^t \mathrm{d}s(\nabla f)^\top(\boldsymbol{x}_{(0)}(t))\boldsymbol{G}_{(0)}(t,s;\boldsymbol{x})\delta\boldsymbol{v}(\boldsymbol{x}_{(0)}(s))\\ &= \int_0^t \mathrm{d}s(\nabla f)^\top(\boldsymbol{x}_{(0)}(t))\boldsymbol{G}_{(0)}(t,t-s;\boldsymbol{x})\delta\boldsymbol{v}(\boldsymbol{x}_{(0)}(t-s)).\end{aligned} \tag{32}$$

Now use $\boldsymbol{G}_{(0)}(t, t-s; \boldsymbol{x}) = \boldsymbol{G}_{(0)}(s, 0; \boldsymbol{x}_{(0)}(t-s))$ and $\boldsymbol{x}_{(0)}(t; \boldsymbol{x}) = \boldsymbol{x}_{(0)}(s; \boldsymbol{x}_{(0)}(t-s; \boldsymbol{x}))$, which hold for any autonomous dynamics. Thus,

$$\delta f(\boldsymbol{x}(t)) = \int_0^t \mathrm{d}s (\boldsymbol{\nabla} f)^\top (\boldsymbol{x}_{(0)}(s; \boldsymbol{x}'))\boldsymbol{G}_{(0)}(s, 0; \boldsymbol{x}')\delta \boldsymbol{v}(\boldsymbol{x}')\Big|_{\boldsymbol{x}'=\boldsymbol{x}_{(0)}(t-s;\boldsymbol{x})}.$$

Appendix B.1.2. Formulation with adjoint variables. We can reformulate this using an 'adjoint variable' $\boldsymbol{\kappa}_f(r, s)$ defined as

$$\boldsymbol{\kappa}_f(r, s; \boldsymbol{x}) := \bar{\boldsymbol{G}}_{(0)}(r, s; \boldsymbol{x})(\boldsymbol{\nabla} f)(\boldsymbol{x}_{(0)}(s; \boldsymbol{x})). \tag{33}$$

Then,

$$\delta f(\boldsymbol{x}(t)) = \int_0^t \mathrm{d}s\, \boldsymbol{\kappa}_f^\top(s, \boldsymbol{x}')\delta \boldsymbol{v}(\boldsymbol{x}')\Big|_{\boldsymbol{x}'=\boldsymbol{x}_{(0)}(t-s;\boldsymbol{x})}, \tag{34}$$

where we further denote $\boldsymbol{\kappa}_f(s) = \boldsymbol{\kappa}_f(0, s)$. Observe that $\boldsymbol{\kappa}_f(r, s; \boldsymbol{x})$ solves the *adjoint equation*

$$\partial_r \boldsymbol{\kappa}_f(r, s) + \left(\frac{\partial \boldsymbol{v}_{(0)}}{\partial \boldsymbol{x}}\right)^\top (\boldsymbol{x}_{(0)}(r))\boldsymbol{\kappa}_f(r, s) = \boldsymbol{0}, \tag{35}$$

backward in time, with the final condition

$$\boldsymbol{\kappa}_f(r, s)|_{r=s} = (\boldsymbol{\nabla} f)(\boldsymbol{x}_{(0)}(s)). \tag{36}$$

The meaning of $\boldsymbol{\kappa}_f(t; \boldsymbol{x})$ is very simple. Note that

$$\frac{\partial \boldsymbol{x}_{(0)}(t; \boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{G}(t, 0; \boldsymbol{x}).$$

Thus,

$$\begin{aligned}\boldsymbol{\kappa}_f^\top(t; \boldsymbol{x}) &= (\boldsymbol{\nabla} f)^\top (\boldsymbol{x}_{(0)}(t; \boldsymbol{x}))\boldsymbol{G}(t, 0; \boldsymbol{x}) \\ &= (\boldsymbol{\nabla} f)^\top (\boldsymbol{x}_{(0)}(t; \boldsymbol{x}))\frac{\partial \boldsymbol{x}_{(0)}(t; \boldsymbol{x})}{\partial \boldsymbol{x}} \\ &= \frac{\partial f(\boldsymbol{x}_{(0)}(t; \boldsymbol{x}))}{\partial \boldsymbol{x}},\end{aligned} \tag{37}$$

by the chain rule. The variable $\boldsymbol{\kappa}_f(t; \boldsymbol{x})$ therefore gives the response of $f(\boldsymbol{x}_{(0)}(t; \boldsymbol{x}))$ to a change of the initial datum $\boldsymbol{x}$.

Appendix B.1.3. Averaging over initial data. If we average (34) over initial data, it becomes

$$\delta\langle f\rangle^t = \int_0^t \mathrm{d}s \langle \boldsymbol{\kappa}_f(s)\delta \boldsymbol{v}\rangle_{(0)}^{t-s}.$$

Here $\langle\cdot\rangle_{(0)}^{t-s}$ denotes that the average over the initial data $\langle\cdot\rangle$ is evolved under the zeroth-order (unperturbed) dynamics for a time $t - s$. In particular, if we take the distribution over initial data to be $P^*_{(0)}(\boldsymbol{x})$, the stationary distribution of the unperturbed dynamics, then

$$\delta\langle f\rangle^t = \int_0^t \mathrm{d}s \langle \boldsymbol{\kappa}_f^\top(s)\delta \boldsymbol{v}\rangle_{(0)}^{*}. \tag{38}$$

The above formula is our main result, *Ruelle's general linear response formula* [5, 6, 8]. Note that our results here for the finite-time response are completely rigorous, without any ergodic-like assumptions on the dynamics, and are a simple application of calculus.

Appendix B.2. The Lea–Allen–Haine approach

Appendix B.2.1. Time-average sensitivity. The procedure considered by Lea *et al* in [10, 11] was instead to apply standard sensitivity analysis to the time-average

$$\bar{f}^\tau(\boldsymbol{x}) = \frac{1}{\tau}\int_0^\tau \mathrm{d}t f(\boldsymbol{x}(t;\boldsymbol{x})) \tag{39}$$

(e.g. see [9]) and then to average over an ensemble $\boldsymbol{x}_n$, $n = 1, \ldots, N$ of initial conditions. For the sake of comparison, we rederive the standard algorithm here.

From (32) we see that

$$\delta\bar{f}^\tau = \frac{1}{\tau}\int_0^\tau \mathrm{d}t \int_0^t \mathrm{d}s (\boldsymbol{\nabla} f)^\top(\boldsymbol{x}_{(0)}(t))\boldsymbol{G}_{(0)}(t,s;\boldsymbol{x})\delta\boldsymbol{v}(\boldsymbol{x}_{(0)}(s)) \tag{40}$$

$$= \frac{1}{\tau}\int_0^\tau \mathrm{d}s \int_s^\tau \mathrm{d}t (\boldsymbol{\nabla} f)^\top(\boldsymbol{x}_{(0)}(t))\boldsymbol{G}_{(0)}(t,s;\boldsymbol{x})\delta\boldsymbol{v}(\boldsymbol{x}_{(0)}(s)) \tag{41}$$

$$= \frac{1}{\tau}\int_0^\tau \mathrm{d}s\, \boldsymbol{\lambda}^\top(s)\delta\boldsymbol{v}(\boldsymbol{x}_{(0)}(s)) \tag{42}$$

with

$$\boldsymbol{\lambda}(s) := \int_s^\tau \mathrm{d}t\, \bar{\boldsymbol{G}}_{(0)}(s,t)(\boldsymbol{\nabla} f)(\boldsymbol{x}_{(0)}(t)).$$

From the properties of the response function given before, it follows that

$$[\partial_s + \boldsymbol{A}^\top(s)]\boldsymbol{\lambda}(s) = -(\boldsymbol{\nabla} f)(\boldsymbol{x}(s))$$

with final datum $\boldsymbol{\lambda}(\tau) = \boldsymbol{0}$.

If we now average as well over initial data, then using

$$\langle\bar{f}^\tau\rangle = \frac{1}{\tau}\int_0^\tau \mathrm{d}t \langle f\rangle^t, \tag{43}$$

we see that the response $\delta\langle\bar{f}^\tau\rangle$ is just the time-average of the response $\delta\langle f\rangle^t$ calculated using Ruelle's formula.

References

[1] Toda M, Kubo R, Saito N and Hashitsume N 1991 *Statistical Mechanics* vol 2, 2nd edn (Berlin: Springer)

[2] Evans D J and Morriss G P 1990 *Statistical Mechanics of Nonequilibrium Liquids* (London: Academic)

[3] Green M S 1954 Markoff random processes and the statistical mechanics of time-dependent phenomena. II. Irreversible processes in fluids *J. Chem. Phys.* **22** 398–413

[4] Kubo R 1957 Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems *J. Phys. Soc. Japan* **12** 570–86

[5] Ruelle D 1997 Differentiation of SRB states *Commun. Math. Phys.* **187** 227–41

[6] Gallavotti G and Ruelle D 1997 SRB states and nonequilibrium statistical mechanics close to equilibrium *Commun. Math. Phys.* **190** 279–85

[7] Ruelle D 1998 Nonequilibrium statistical mechanics near equilibrium: computing higher order terms *Nonlinearity* **11** 5–18

[8] Ruelle D 1998 General linear response formula in statistical mechanics, and the fluctuation–dissipation theorem far from equilibrium *Phys. Lett.* A **245** 220–4

[9] Errico R M 1997 What is an adjoint model? *Bull. Am. Meteorol. Soc.* **78** 2576–91

[10] Lea D J, Allen M R and Haine T W N 2000 Sensitivity analysis of the climate of a chaotic system *Tellus* A **52** 523–32

[11] Lea D J, Allen M R, Haine T W N and Hansen J 2002 Sensitivity analysis of the climate of a chaotic ocean circulation model *Q. J. R. Meteorol. Soc.* **128** 2587–606

[12] Lorenz E N 1963 Deterministic nonperiodic flow *J. Atmos. Sci.* **20** 130–41

[13] Eckmann J-P and Ruelle D 1985 Ergodic theory of chaos and strange attractors *Rev. Mod. Phys.* **57** 617–56

[14] Gallavotti G and Cohen E G D 1995 Dynamical ensembles in stationary states *J. Stat. Phys.* **80** 931–70

[15] Gallavotti G and Cohen E G D 1995 Dynamical ensembles in nonequilibrium statistical mechanics *Phys. Rev. Lett.* **74** 2694–7

[16] Gallavotti G 1998 Chaotic dynamics, fluctuations, non-equilibrium ensembles *Chaos* **8** 384–92

[17] Ruelle D 1999 Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics *J. Stat. Phys.* **95** 393–468

[18] Chernov N, Eyink G, Lebowitz J L and Sinai Y 1993 Steady-state electrical conduction in the periodic Lorentz gas *Commun. Math. Phys.* **54** 569–601

[19] Chernov N, Eyink G, Lebowitz J L and Sinai Y 1993 Derivation of Ohm's law in a deterministic mechanical model *Phys. Rev. Lett.* **70** 2209–12

[20] Eyink G L, Lebowitz J L and Spohn H 1990 Microscopic origin of hydrodynamic behavior: entropy production and the steady state *CHAOS/XAOC. Soviet–American Perspectives on Nonlinear Science* ed D K Campbell (New York: American Institute of Physics)

[21] Gnedenko B V and Kolmogorov A N 1968 *Limit Distributions for Sums of Independent Random Variables.* (Reading, MA: Addison-Wesley)

[22] Chandrasekhar S 1943 Stochastic problems in physics and astronomy *Rev. Mod. Phys.* **15** 1–89

[23] Mandelbrot B 1963 The variation of certain speculative prices *J. Business* **36** 394–419

[24] Mantegna R N and Stanley H E 1995 Scaling behaviour in the dynamics of an economic index *Nature* **376** 46–9

[25] Bouchaud J-P and Georges A 1990 Anomalous diffusion in disordered media: statistical mechanics, models and physical applications *Phys. Rep.* **195** 127–293

[26] Umeno K 1998 Superposition of chaotic process with convergence to Levy's stable law *Phys. Rev.* E **58** 2644–7

[27] Krauskopf B and Osinga H M 2004 The Lorenz manifold as a collection of geodesic level sets *Nonlinearity* **17** C1–C6

[28] Sparrow C 1982 The Lorenz equations: bifurcations, chaos, and strange attractors *Applied Mathematical Sciences* vol 41 (New York: Springer)

[29] Tucker W 1999 The Lorenz attractor exists *C. R. Acad. Sci. Paris Séries I Math.* **328** 1197–202

[30] Tucker W 2002 A rigorous ODE solver and Smale's 14th problem *Found. Comput. Math.* **2** 53–117

[31] Rössler O E 1976 An equation for continuous chaos *Phys. Lett.* A **57** 397–8

[32] McShane E J 1989 The calculus of variations from the beginning through optimal control theory *SIAM J. Control Optim.* **27** 916–39

[33] Alder B J and Wainwright T E 1970 Decay of the velocity autocorrelation function *Phys. Rev.* A **1** 18–21

[34] Ernst M H, Cichocki B, Dorfman J R, Sharma J and van Beijeren H 1978 Kinetic theory of nonlinear viscous flow in two and three dimensions *J. Stat. Phys.* **18** 237–70