

# Probability of noise and rate-induced tipping

Paul Ritchie\* and Jan Sieber†

*Centre for Systems, Dynamics and Control, College of Engineering,  
Mathematics and Physical Sciences, Harrison Building,  
University of Exeter, Exeter, EX4 4QF, United Kingdom*

(Dated: April 13, 2017)

We propose an approximation for the probability of tipping when the speed of parameter change and additive white noise interact to cause tipping. Our approximation is valid for small to moderate drift speeds and helps to estimate the probability of false positives and false negatives in early-warning indicators in the case of rate- and noise-induced tipping. We illustrate our approximation on a prototypical model for rate-induced tipping with additive noise using Monte-Carlo simulations. The formula can be extended to close encounters of rate-induced tipping and is otherwise applicable to other forms of tipping.

We also provide an asymptotic formula for the critical ramp speed of the parameter in the absence of noise for a general class of systems undergoing rate-induced tipping.

## I. INTRODUCTION

The notion of tipping describes the phenomenon observed in science, where gradual changes to input levels cause a sudden (in practice possibly catastrophic) change in the output. Examples of tipping in science include: Arctic sea ice melting [1], degradation of coral reefs [2], dieback of tropical forest and savanna to a treeless state [3] and financial market crashes [4].

Recently, Ashwin *et al.* [5] attempted to classify the underlying mechanisms behind any observed tipping event as an example of either bifurcation-, noise- or rate-induced tipping. The case of a slow passage through a bifurcation (often a saddle-node), causing a loss of stability and therefore an abrupt transition to an alternative stable state [6], is called *bifurcation-induced tipping*. In contrast, random (rare) jumps between attractors of an underlying deterministic system due to fluctuations is classified as *noise-induced tipping*. *Rate-induced tipping* occurs when a system fails to track the continuously changing quasi-steady state [5] because the parameter drift speed exceeds a certain critical rate. For more general definitions and properties of bifurcation- or rate-induced tipping we refer to Ashwin *et al.* [7].

A research area related to tipping is the study of generic early-warning indicators [8]. Increase of autocorrelation and variance in output time series are two statistical indicators which are based on the phenomenon commonly known as ‘critical slowing down’ as a system parameter approaches a bifurcation value [9]. The idea is that far from a bifurcation, the state of the system behaves like an overdamped particle in a slowly softening potential well [10]. If a small perturbation is made to the particle there will be a fast recovery back to the equilibrium [11]. However, as the bifurcation is approached, the well softens and the recovery from a small perturbation will be slower such that one generically observes

an increase in the autocorrelation and variance in output time series [12]. On the other hand, for purely noise-induced transitions no bifurcation point is approached and therefore there is debate into the usefulness of the early-warning signals for this type of tipping [13, 14]. Rate-induced tipping does not involve a loss of stability [15] and therefore Ashwin *et al.* [5] commented that there is no reason to suggest the early-warning indicators should be present. However, it has been shown that for a prototypical model for rate-induced tipping the autocorrelation and variance increase before the closest encounter with the critical rate occurs [10].

In the study of palaeoclimate records [16, 17] early-warning indicators have been tested on events in the past when tipping has been known to occur. However, testing the early-warning indicators against historical examples is susceptible to statistical mistakes as one selects data conditioned on the system having tipped [18].

A natural progression will be to use the early-warning indicators to try and predict future tipping events. Though, this raises such questions as, if we were to observe an increase in both the autocorrelation and variance of a time series does this mean that the system will tip? Boettiger and Hastings [19] show there is an increased rate of false positives in early-warning indicators for simulated systems that experience transitions purely by chance. Furthermore, Drake [20] suggests that stochastic switching can be anticipated but argues that any statistics to be used as early-warning requires decision theory to balance the strength of evidence against the cost and benefits of early-warnings and false positives.

This paper provides a generic approximation for the probability of a prototypical model for rate-induced tipping with additive white noise. It is structured as follows: Section II gives a general asymptotic approximation for the critical rate for a class of deterministic systems with rate-induced tipping. Section III derives the approximation for the probability of a noise-induced escape during ramp of a system parameter that does not quite reach the critical rate. In Section IV, we illustrate the general ap-

\* pdlr201@exeter.ac.uk

† J.Sieber@exeter.ac.uk

proximation result with the prototype model introduced by Ashwin *et al.* [5], which we then systematically study in dependence of its parameter in Section V. Finally, Section VI the limits of our approximations and how further developments can address these.

## II. RATE-INDUCED TIPPING IN SYSTEMS WITH A RAMPED PARAMETER

A general scenario for the phenomenon of rate-induced tipping was considered by Ashwin *et al.* [5, 7]. Assume that a parameter  $\lambda$  corresponds to a shift of the coordinate system:

$$\dot{x} = f(x + b\lambda), \quad x(t) \in \mathbb{R}^n, \quad b \in \mathbb{R}^n, \quad \lambda \in \mathbb{R}, \quad (1)$$

where the vector  $b$  is the direction of the shift and  $\lambda$  is the (scalar) amount. For each fixed  $\lambda$  the stability of, for example, equilibria of (1) is identical. However, when  $\lambda$  is time-dependent, then there can be *critical rates* [21] of change of  $\lambda$ .

*a. Linear shift* The simplest example discussed in Ashwin *et al.* [5] is a linear parameter shift, that is,  $\lambda = r_{\text{lin}}t$  (with  $r_{\text{lin}} > 0$  constant). One of the cases studied in Ashwin *et al.* [5] was assuming that the system in co-moving coordinates

$$\dot{y} = f(y) + r_{\text{lin}}b \quad (\text{where } y = x + br_{\text{lin}}t) \quad (2)$$

has a saddle-node bifurcation at  $r_{\text{lin}} = r_0 > 0$ ,  $y = y_0$  with a stable branch  $y^{(s)}[r_{\text{lin}}]$  and an unstable branch  $y^{(u)}[r_{\text{lin}}]$  of equilibria emerging for  $r_{\text{lin}} \in [0, r_0)$ . (We will be using square brackets to denote branches of equilibria to avoid confusion with time dependence.) These equilibria for  $y$  correspond to stable and unstable invariant lines  $x^{(s)}(t) = y^{(s)}[r_{\text{lin}}] - br_{\text{lin}}t$ ,  $x^{(u)}(t) = y^{(u)}[r_{\text{lin}}] - br_{\text{lin}}t$  of the original system (1) for  $r_{\text{lin}} \in [0, r_0)$ . For  $r_{\text{lin}} < r_0$ , all initial conditions  $x(0)$  near  $y^{(s)}[r_{\text{lin}}]$  follow  $x^{(s)}(t)$  for all  $t > 0$ , while for  $r_{\text{lin}} > r_0$  this invariant line no longer exists such that the rate  $r_0$  is critical. This scenario corresponds to a saddle-node bifurcation in the co-moving coordinates (2) using  $y$ . Increasing  $r_{\text{lin}}$  gradually corresponds to a slow passage through a saddle-node bifurcation.

*b. Ramped shift* A more complex scenario is the case where  $\lambda$  is “ramped up”, that is,  $\lambda \rightarrow 0$  for  $t \rightarrow -\infty$ ,  $\lambda \rightarrow \lambda_{\text{max}}$  for  $t \rightarrow +\infty$ , and  $\dot{\lambda}(t) > 0$  for all  $t$ . A prototype system for this ramping scenario was studied in [5, 10], and is used in Section IV for illustration. For the one-dimensional case Ashwin *et al.* [7] gave topological criteria (and a general definition) for rate-induced tipping with a ramped parameter  $\lambda$ . For the general case (1) we assume that the change of the ramp in  $\lambda$  is itself given by a scalar differential equation. Define

$$\begin{aligned} r &= \max\{\dot{\lambda}(t) : t \in \mathbb{R}\} && \text{(maximal ramp speed)} \\ \epsilon &= r/\lambda_{\text{max}} && \text{(sharpness of ramp)} \\ \mu(t) &= \lambda(t)/\lambda_{\text{max}} && (\lambda \text{ normalized to } [0, 1]), \end{aligned}$$

then, assuming  $\dot{\lambda}$  is bounded and using the new parameters  $r$  and  $\epsilon$ ,  $x$  and  $\mu$  are the solution of an autonomous extended system:

$$\dot{x} = f\left(x + b\frac{r}{\epsilon}\mu\right), \quad (3)$$

$$\dot{\mu} = \epsilon\Gamma(\mu), \quad (4)$$

where  $\Gamma$  normalizes  $\dot{\lambda}$  to  $[0, 1]$ . Since  $\dot{\lambda}$  is always positive,  $\Gamma$  satisfies the following properties

$$\begin{aligned} \Gamma(0) = \Gamma(1) = 0, \quad \max\{\Gamma(\mu) : \mu \in [0, 1]\} = 1, \quad \text{and} \\ \Gamma(\mu) > 0 \text{ for all } \mu \in (0, 1). \end{aligned} \quad (5)$$

Let us also assume that  $\lambda$  approaches its limits at an exponential rate such that  $\Gamma'(0) > 0$ ,  $\Gamma'(1) < 0$ , and that  $\Gamma$  is only equal to 1 in a single point  $\mu_{\text{crit}} \in (0, 1)$  and that  $\Gamma''(\mu_{\text{crit}}) < 0$ .

If system (2) has a saddle-node bifurcation at  $y = y_0$ ,  $r := r_{\text{lin}} = r_0$ , connecting a stable branch  $y^{(s)}[r]$  of equilibria of (2) and a branch  $y^{(u)}[r]$  with a single degree of instability for  $r \in [0, r_0]$ , then we can make the following statement about the existence of a critical rate  $r_c(\epsilon)$  for sufficiently small  $\epsilon$ .

The system (3)–(4) has (at least) 4 equilibria:

- $x_{\text{eq}}^{(s,0)} := y^{(s)}[r] \Big|_{r=0}$ ,  $\mu = 0$  with one unstable direction,
- $x_{\text{eq}}^{(u,0)} := y^{(u)}[r] \Big|_{r=0}$ ,  $\mu = 0$  with two unstable directions,
- $x_{\text{eq}}^{(s,1)} := y^{(s)}[r] \Big|_{r=0} - br/\epsilon$ ,  $\mu = 1$  (stable),
- $x_{\text{eq}}^{(u,1)} := y^{(u)}[r] \Big|_{r=0} - br/\epsilon$ ,  $\mu = 1$  with one unstable direction.

For sufficiently small  $\epsilon$  there are three possible scenarios for system (3)–(4) depending on  $r$  and a critical rate  $r_c(\epsilon)$ , illustrated in Figure 1.

**Tracking:**  $r < r_c(\epsilon)$ . There is a connecting orbit from  $(x_{\text{eq}}^{(s,0)}, 0)$  to  $(x_{\text{eq}}^{(s,1)}, 1)$ . In this case, solutions  $(x(t), \mu(t))$  starting close to  $(x_{\text{eq}}^{(s,0)}, 0)$  stay close to  $(x_{\text{eq}}^{(s,0)} - br\mu(t)/\epsilon, \mu(t))$  for all  $t$  (the distance goes to 0 as  $\epsilon \rightarrow 0$ ), see Figure 1a.

**Critical:**  $r = r_c(\epsilon) = r_0 + O(\epsilon)$ . There is a saddle-to-saddle connection from  $(x_{\text{eq}}^{(s,0)}, 0)$  to  $(x_{\text{eq}}^{(u,1)}, 1)$  in system (3)–(4). The first-order expansion for  $r_c$  in  $\epsilon$  is

$$r_c(\epsilon) = r_0 + \epsilon \sqrt{\frac{-r_0 \Gamma''(\mu_{\text{crit}})}{2a_0 a_2}} + O(\epsilon^2), \quad (6)$$

where  $a_0 = w_0^T b$ ,  $a_2 = \frac{1}{2} w_0^T \partial^2 f(y_0) v_0^2$ , and  $w_0$  and  $v_0$  are the left and right nullvectors of  $\partial f(y_0)$ , scaled such that  $w_0^T v_0 = 1$  and  $a_0 a_2 > 0$ . The coefficients  $a_0$  and  $a_2$  are

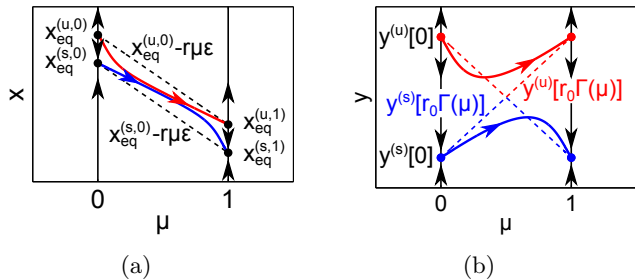


FIG. 1: (a) Phase plane of system (3)–(4) for the scenario of tracking,  $r < r_c(\epsilon)$ . Black dashed lines are the stable (lower) and unstable (upper) branches of equilibria in the limit  $\epsilon = 0$ . Solid blue and red curves represent connecting orbits between  $(x, \mu) = (x_{\text{eq}}^{(s,0)}, 0)$  and  $(x_{\text{eq}}^{(s,1)}, 1)$  and  $(x_{\text{eq}}^{(u,0)}, 0)$  to  $(x_{\text{eq}}^{(u,1)}, 1)$  respectively. (b) Phase plane of system (8)–(9) for  $r = r_0$ , tracking scenario for  $\epsilon > 0$ . The blue (lower) dashed line displays the branch of stable equilibria and the red (upper) dashed line is the branch of unstable equilibria in the limit  $\epsilon = 0$ . Connecting orbits between  $(y, \mu) = (y^{(s)}[0], 0)$  and  $(y^{(s)}[0], 1)$  and  $(y^{(u)}[0], 0)$  and  $(y^{(u)}[0], 1)$  given by solid blue and red curves respectively. Parameters:  $b = 1$ ,  $\epsilon = 0.21$ .

the expansion coefficients when one inserts  $y = y_0 + v_0 z$  and  $r_{\text{lin}} = r\Gamma(\mu) - r_0$  into (2), applies  $w_0^T$ , and truncates to second-order terms:

$$\dot{z} = a_0(r\Gamma(\mu) - r_0) + a_2 z^2 + O(z^3). \quad (7)$$

See Appendix A for details of the derivation of the first-order expansion for  $r_c$  given by equation (6).

**Escape:**  $r > r_c(\epsilon)$ . There are initial conditions for (3)–(4) arbitrarily close to  $(x_{\text{eq}}^{(s,0)}, 0)$  that escape, following the unstable manifold of  $(x_{\text{eq}}^{(u,1)}, 1)$ .

The expression for  $r_c(\epsilon)$  shows that in the limit  $\max \dot{\lambda} \ll \lambda_{\text{max}}$  ( $\epsilon \ll 1$ , long “gentle” ramps) the critical rate  $r_c(\epsilon)$  for the ramp approaches the critical rate  $r_0$  for the linear shift from above.

The expansion (6) for the saddle-connection is determined entirely by quantities close to the saddle-node, because  $x(t)$  is  $\epsilon$ -close to  $y^{(s)}[r\Gamma(\mu)]$  for  $\mu < \mu_{\text{crit}}$ , and it is  $\epsilon$ -close to  $y^{(u)}[r\Gamma(\mu)]$  for  $\mu > \mu_{\text{crit}}$  (recall that  $\Gamma$  equals 1 only for  $\mu = \mu_{\text{crit}}$ ).

Figure 1b illustrates the phase space for the example from Section IV in co-moving coordinates (for scalar  $y$ )

$$\begin{aligned} \dot{y} &= f(y) + br\Gamma(\mu), \\ \dot{\mu} &= \epsilon\Gamma(\mu) \end{aligned} \quad (8)$$

$$\dot{\mu} = \epsilon\Gamma(\mu) \quad (9)$$

with  $r = r_0$ . This is an illustration of the tracking scenario ( $r < r_c(\epsilon)$ ) and so there is a connecting orbit between  $(y, \mu) = (y^{(s)}[0], 0)$  and  $(y^{(s)}[0], 1)$ . The distance of the connecting orbit to  $y^{(s)}[r_0\Gamma(\mu)] \rightarrow 0$  as  $\epsilon \rightarrow 0$ .

The branches of equilibria  $y^{(s)}[r\Gamma(\mu)]$  and  $y^{(u)}[r\Gamma(\mu)]$  change their arrangement depending on the value of  $r$  in

relation to  $r_0$ . For  $r < r_0$  there exist a continuous branch of stable (unstable) equilibria connecting  $y^{(s)}[0]$  ( $y^{(u)}[0]$ ) between  $\mu = 0$  and  $\mu = 1$ . For  $r = r_0$  the two branches meet as depicted by Figure 1b. The connections then break up for  $r > r_0$  such that two separate saddle-node bifurcations are formed. Appendix B contains illustrations for the phase space for other rates.

### III. NOISE-INDUCED ESCAPE DURING RAMP NEAR BUT BELOW CRITICAL RATE

We consider the effect of additive noise for the scalar setting. Then, system (3)–(4) changes into a scalar stochastic differential equation (SDE) for a random variable  $X_t$

$$dX_t = f\left(X_t + \frac{r}{\epsilon}\mu\right) dt + \sqrt{2D}dW_t, \quad (10)$$

$$\dot{\mu} = \epsilon\Gamma(\mu), \quad (11)$$

where  $W_t$  is standard Brownian motion, the intensity of the noise is given by  $\sqrt{2D}$ , and  $D$  is a constant diffusion coefficient. We assume that the deterministic part is as described in Section II. The deterministic part in (10) corresponds to a choice of  $b$  equal to 1 in the general equation (3). Setting  $Y_t = X_t + r\mu/\epsilon$  gives

$$dY_t = (f(Y_t) + r\Gamma(\mu(t)))dt + \sqrt{2D}dW_t \quad (12)$$

where  $\dot{y} = f(y) + r\Gamma(\mu)$  follows the scenario from Section II when treating  $\mu$  as a parameter. That is, for  $r = r_0$ , the system touches a saddle-node non-transversally when  $\mu$  crosses  $\mu_{\text{crit}}$ . As introduced in Section II, the  $y^{(s)}$  and  $y^{(u)}$  are the stable and unstable branches. They are arranged such that  $y^{(s)}[r\Gamma(\mu)] < y^{(u)}[r\Gamma(\mu)]$ , as shown in Figure 1b, when  $r\Gamma(\mu) \in [0, r_0]$  (consistent with our setup in Section II and the example in Section IV). This means that

$$x_{\text{eq}}^{(s,0)} = y^{(s)}[r\Gamma(\mu)] \Big|_{r=0} < x_{\text{eq}}^{(u,0)} = y^{(u)}[r\Gamma(\mu)] \Big|_{r=0}$$

are equilibria of the deterministic part of (10) combined with  $\dot{\mu} = \epsilon\Gamma(\mu)$  (identical to (4) with the same assumptions (5) on  $\Gamma$ ). Expressions below also use the potential  $U_r$  for the deterministic part of (12):

$$U_r(y) = -\int f(y)dy - r\Gamma y.$$

*Stationary case with noise* (Equation (10) with  $r = 0$ , which is independent of  $\mu$ ) For  $D > 0$  there will be a fixed escape rate  $\kappa$  from the basin of attraction of  $x_{\text{eq}}^{(s,0)}$  across  $x_{\text{eq}}^{(u,0)}$ , approximated by Kramers’ escape rate:

$$\kappa \approx \frac{\sqrt{\alpha\beta}}{2\pi} \exp\left(-\frac{\Delta U_0}{D}\right), \quad (13)$$

where  $\alpha = \partial_{yy}U_0(x_{\text{eq}}^{(s,0)})$ ,  $\beta = -\partial_{yy}U_0(x_{\text{eq}}^{(u,0)})$  and  $\Delta U_r = U_r(y^{(u)}[r]) - U_r(y^{(s)}[r])$  ( $\Delta U_0$  is then  $\Delta U_r$  at  $r = 0$ ). The approximation (13) is accurate for  $D \ll \Delta U_0$ .

### A. First-order approximations of non-stationary Fokker-Planck Equation (FPE)

We study the first-order deviation from the quasi-stationary escape rate for  $0 < \epsilon \ll 1$ . For the following section we fix a specific trajectory for  $\mu(t)$  in  $\dot{\mu} = \epsilon\Gamma(\mu)$  by choosing the initial condition such that

$$\mu = \mu_{\text{crit}} \quad \text{for } t = 0$$

(remember that  $\Gamma(\mu_{\text{crit}}) = 1$  is the unique maximum of  $\Gamma$ ). We are interested in the parameter range where the diffusion coefficient  $D$  and the maximal ramp speed  $r$  satisfy

$$D = O(\epsilon^{3/2}), \quad r - r_c(\epsilon) = O(\epsilon) < 0.$$

Thus, the maximal ramp speed  $r$  is close to, but below its critical value  $r_c(\epsilon)$  given in (6). The scaling of the diffusion coefficient arises naturally from the change of coordinates to (7) with white noise of variance  $2D$ ; see Appendix A. For  $D \ll \epsilon^{3/2}$  we have a small-noise limit for  $\epsilon \rightarrow 0$  for the probability of escape, governed by Kramers' escape rate (13). For  $D \gg \epsilon^{3/2}$  the probability of escape is dominated by noise-induced escape far away from the tipping. For  $D \sim \epsilon^{3/2}$  all coefficients in the non-dimensionalized system are of order unity such that the time-dependence and the noise effects are in non-trivial balance for all small  $\epsilon$ . The scaling implies in particular that

$$0 < D \sim \Delta U_r \ll \Delta U_0. \quad (14)$$

**Remarks:** (a) Condition (14) means that the escape rate is small before and after the ramp ( $t \rightarrow \pm\infty$ ,  $\mu(t)$  far from  $\mu_{\text{crit}}$ ) such that the Kramers approximation (13) for the escape rate is applicable for all times outside of an interval  $[t_0, T_{\text{end}}]$  around  $t = 0$ .

(b) Condition (14) also means that the maximal ramp speed  $r$  is sufficiently large such that approximation (13) is no longer true at the maximum speed of the ramp ( $t = 0$ ,  $\mu = \mu_{\text{crit}}$ ), but it is still less than the saddle-node rate  $r_0$  (note that  $\Delta U_{r_0} = 0$ ), the limit of the critical rate  $r_c(\epsilon)$  for  $\epsilon \rightarrow 0$ . Thus, without noise ( $D = 0$ ), there is a connection from  $(x_{\text{eq}}^{(s,0)}, 0)$  to  $(x_{\text{eq}}^{(s,1)}, 1)$  (the case of *tracking* in Section II). Let us pick one time profile  $\tilde{x}(t)$  on this connecting orbit.

We consider a starting position  $(x, t) = (x_0, t_0)$  and an end position at  $(x, t) = (x_T, T_{\text{end}})$ , and a strip  $S_\delta$  of width  $2\delta$  and length  $T_{\text{end}} - t_0$  around  $\tilde{x}$ , see Figure 2.

The Fokker-Planck equation

$$\frac{\partial P}{\partial t} = D \frac{\partial^2 P}{\partial x^2} - \frac{\partial}{\partial x} \left( f(x + r\mu(t)/\epsilon) P \right), \quad (15)$$

describes the time evolution of the probability density  $P(x, t)$  of the random variable  $X_t$ , governed by (10). If we impose Dirichlet boundary conditions,

$$0 = P(\tilde{x}(t) + \delta, t), \quad (16)$$

$$0 = P(\tilde{x}(t) - \delta, t), \quad (17)$$

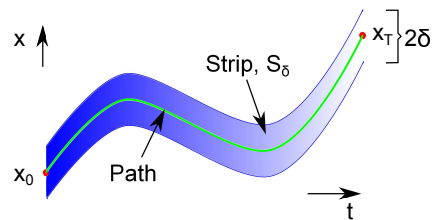


FIG. 2: Sketch of the path  $\tilde{x}(t)$  (green curve) and the surrounding strip  $S_\delta$  of width  $2\delta$  (shaded blue). The probability density of a realization passing through  $(x, t)$  (always staying within  $S_\delta$ ) is  $P(x, t)$ .

then  $\int_a^e P(x, t) dx$  is the probability that the solution of (10), starting at  $t_0$  with probability density  $P(\cdot, t_0)$ , is in  $[a, e]$  at time  $t$  and has never left the strip  $S_\delta$ .

Consequently, the overall escape probability from the strip  $S_\delta$  of width  $2\delta$  around the path  $\tilde{x}$  during time interval  $[t_0, T_{\text{end}}]$  equals  $1 - \int_{-\delta}^{\delta} P(\tilde{x}(T_{\text{end}}) + x, T_{\text{end}}) dx$ .

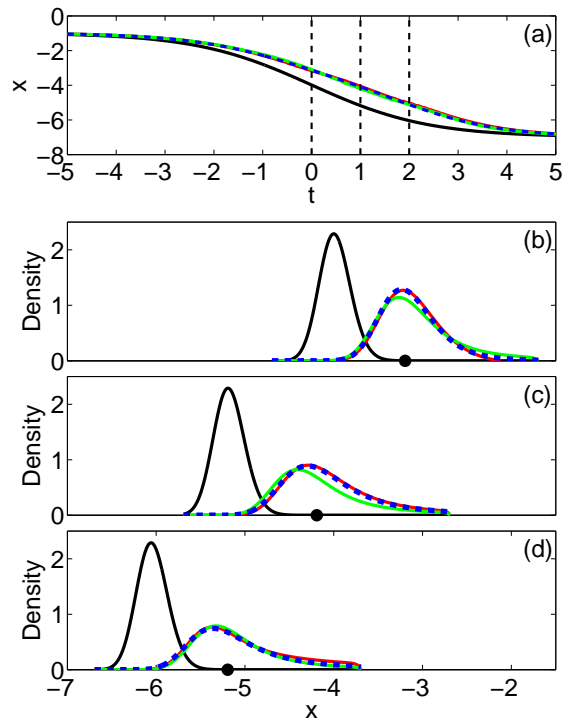


FIG. 3: Comparison of the single- (green (light gray)) and three-mode (red (dark gray)) approximations with the density from simulations (blue dashed) and stationary density (black) for  $\epsilon = 0.21$ ,  $r = 1.26$ . Panel (a) provides the time profile for the location of the mean of each distribution. Vertical dashed lines indicate the times for which the densities are given in the remaining panels, namely (b)  $t = 0$ , (c)  $t = 1$ , and (d)  $t = 2$ . Black dot on  $x$ -axis (panels (b)-(d)) corresponds to location of deterministic trajectory starting at  $x_0 = -1$  at  $t = -10$ . Parameters: Noise level  $D = 0.06$ , width of strip  $S_\delta$   $2\delta = 3$ .

Figure 3 illustrates the shape of this probability density  $P(x, t)$  along the strip  $S_\delta$  for a ramp speed  $r < r_c(\epsilon)$ . The figure uses parameters from the specific example introduced in Section IV. The first moment of  $P(x, t)$  (the mean) is shown in panel (a) of Figure 3. Panels (b)–(d) show the profile of  $P(x, t)$  for selected times  $t$ . The numerical solution of (15)–(16) is shown as a blue (dashed) curve. The other curves are the approximations described below.

*a. Uncentered quasi-stationary density* The crudest approximation assumes that the density is approximately stationary throughout the ramp. This implies that  $\partial_t P$  is small in (15). Replacing  $\partial_t P$  with zero in (15) and imposing a Dirichlet boundary condition on the right end ( $P(\tilde{x}(t) + \delta, t) = 0$ ), the solution  $P$  of (15) has the form

$$P_*(x, t) = P_0(t) \int_x^{\tilde{x}(t)+\delta} \exp\left[\frac{U(x', t) - U(x, t)}{D}\right] dx', \quad (18)$$

where  $\partial_x U(x, t) = -f(x + r\mu(t)/\epsilon)$  for each fixed  $t$ . The spatial shape of  $P_*(\cdot, t)$  is nearly unchanged, only shifted by  $r\mu(t)/\epsilon$  for different times  $t$ . As the density has nearly constant shape for all  $t$ , the escape rate is nearly constant in time as well (hence, it is equal to  $\kappa$  by remark (a)). This escape rate determines the normalization constant  $P_0(t)$ :  $P_0(t) \approx P_\infty(1 - (t - t_0)\kappa)$ , where  $P_\infty$  is such that the initial density  $P_*(x, t_0)$  has a unit integral. This approximation, shown in black in Figure 3 for the example from Section IV, does not catch the effect of a non-zero  $r$ : it is (nearly) independent of  $r$  (becoming independent of  $r$  in the limit  $\delta \rightarrow \infty$ ). The density  $P_*(x, t)$  is centered at  $x_{\text{eq}}^{(s,0)} - r\mu(t)/\epsilon$ , which is visibly smaller than  $\tilde{x}(t)$  (the location of the deterministic trajectory, highlighted by the black dot on the  $x$ -axis in panels (b)–(d) of Figure 3).

*b. Instantaneous eigenmodes of FPE, centered at  $\tilde{x}$*  The instantaneous eigenmode expansion for the linear operator of the Fokker-Planck equation (15) follows an approach similar to that presented in Risken and Frank [22], Zhang *et al.* [23] but for a time dependent deterministic part  $f(x + r\mu(t)/\epsilon)$  instead of a time-independent  $f(x)$ . Figure 3 shows the single-mode approximation ( $n = 1$ ) in green (light gray) and the three-mode approximation ( $n = 3$ ) of the probability density in red (dark gray).

We first change to a co-moving coordinate system  $x = \tilde{x}(t) + y$  with respect to  $y$ , such that, the SDE (10) has the form

$$\begin{aligned} dy &= -yg(y, t)dt + \sqrt{2D}dW_t, \quad \text{where} \quad (19) \\ g(y, t) &= -\int_0^1 f'(sy + \tilde{y}(t))ds \quad \text{with} \\ \tilde{y}(t) &= y^{(s)}[r\Gamma(\mu(t))] - \epsilon \frac{r\Gamma'(\mu(t))\Gamma(\mu(t))}{f'(y^{(s)}[r\Gamma(\mu(t))])^2} + O(\epsilon^2) \end{aligned}$$

(recall that  $y^{(s)}[r\Gamma(\mu)]$  is the stable equilibrium of  $\dot{y} = f(y) + r\Gamma(\mu)$  with fixed  $\mu$ ). For sufficiently small  $\epsilon$  and  $r < r_0$  we have that in the moving coordinates  $y$  the

equilibrium  $y = 0$  is stable without noise for each fixed  $t$  (by the stability assumption on  $y^{(s)}[r\Gamma(\mu(t))]$  and because  $r < r_0$ ). That is, for  $t \in [t_0, T_{\text{end}}]$

$$g(0, t) = -f'(\tilde{y}(t)) = -f'(y^{(s)}[r\Gamma(\mu(t))]) + O(\epsilon) > 0.$$

The Fokker-Planck equation for the density  $P(y, t)$  over  $y$  operates then on the fixed domain  $[-\delta, \delta]$ :

$$\frac{\partial P}{\partial t} = D \frac{\partial^2 P}{\partial y^2} + \frac{\partial}{\partial y} (yg(y, t)P) =: A(t)P \quad (20)$$

with Dirichlet boundary conditions  $P(-\delta, t) = P(\delta, t) = 0$ . The operator  $A(t)$  is self-adjoint with respect to the scalar product

$$\langle w, v \rangle_t = \int_{-\delta}^{\delta} w(y)v(y) \exp\left(\frac{U(y, t)}{D}\right) dy, \quad (21)$$

where  $U$  is the potential corresponding to the drift  $-yg(y, t)$  ( $\partial_y U(y, t) = yg(y, t)$ ). Since  $g(y, t) > 0$ , this effective potential  $U(y, t)$  has a critical point at  $y = 0$ . For each fixed  $t \in [t_0, T_{\text{end}}]$ , the spectrum of  $A(t)$ , shown later in Figure 5 for the specific example considered in Section IV, consists of eigenvalues  $\gamma_i(t)$  with eigenfunctions  $v_i(y, t)$ :

$$\gamma_i v_i = D \frac{\partial^2 v_i}{\partial y^2} + \frac{\partial}{\partial y} [yg(y, t)v_i] = A(t)v_i. \quad (22)$$

The eigenfunctions  $v_i(\cdot, t)$  (called *instantaneous* modes as they are time-dependent) form an orthonormal basis of  $L^2$  with respect to  $\langle \cdot, \cdot \rangle_t$ . Thus, we can expand the solution  $P(y, t)$  of (20) as a linear combination of the instantaneous eigenmodes  $v_i$ :

$$P(y, t) = \sum_{i=1}^{\infty} a_i(t)v_i(y, t), \quad (23)$$

where the  $a_i(t)$  are scalar coefficients at each time  $t$ .

Inserting the expansion (23) into equation (20), applying  $\langle v_k, \cdot \rangle_t$ , and truncating at a finite  $n$  gives [24]:

$$\begin{aligned} \dot{a}_k &= \gamma_k(t)a_k(t) - \sum_{i=1}^n \langle v_k(t), \dot{v}_i(t) \rangle_t a_i(t), \quad (24) \\ a_i(0) &= \langle v_i(\cdot, 0), P(\cdot, 0) \rangle_0, \end{aligned}$$

where the coupling coefficients  $\langle v_k(t), \dot{v}_i(t) \rangle_t$  are of order  $\epsilon/|\gamma_k(t) - \gamma_i(t)|$  for  $i \neq k$ . Thus, the sum is convergent and the truncated solution

$$P_n(x, t) = \sum_{i=1}^n a_i(t)v_i(x, t) \quad (25)$$

converges to  $P$  for  $n \rightarrow \infty$  and for  $\epsilon \rightarrow 0$  ( $-\gamma_k = O(k^2)$  for positive  $D$  (see Figure 5a for the specific example), so, in particular,  $P_n - P = O(\epsilon)$ ).

Since the initial  $t_0$  is such that  $\mu(t_0)$  is still close to 0,  $\gamma_1(t_0)$  will be very close to zero by remark (a) that escape is unlikely outside of the ramping time interval. Remark (a) also implies that the coefficients of the initial value  $P(\cdot, t_0)$ ,  $a_i(t_0)$ , are close to zero for  $i > 1$ .

Figure 3 illustrates that the truncation error for small  $n$  occurs in the tails of the distribution. For example, for the single-mode approximation with  $n = 1$  the equation (24) simplifies to

$$\dot{a}_1 = [\gamma_1 - \langle v_1, \dot{v}_1 \rangle_t] a_1.$$

Ignoring the  $O(\epsilon)$  term  $\langle v_1, \dot{v}_1 \rangle_t$ , the single-mode approximation results in an approximate solution

$$P_1(y, t) = \exp\left(\int_{t_0}^{T_{\text{end}}} \gamma_1 dt\right) v_1(y, t).$$

Thus, truncation at  $n = 1$  assumes that the density instantaneously adjusts its shape to the shape of the effective potential well  $U(y, t) = -\int yg(y, t)dy$  at every time  $t$ .

For each particular truncation  $n$ , the probability  $\mathbb{P}_M$  for the trajectory of a realization to not remain within the strip  $S_\delta$  is approximately

$$\mathbb{P}_M = 1 - \int_{-\delta}^{\delta} P_n(y, T_{\text{end}}) dy. \quad (26)$$

## B. Perturbation approximation of the dominant eigenvalue

The dominant eigenvalue  $\gamma_1(t)$  and eigenfunction  $v_1$  can be approximated via a linear perturbation analysis from the small-noise limit ( $D \rightarrow 0$ ). Hence, we can approximate the dominant term in  $\mathbb{P}_M$  for the truncation  $n = 1$ , which is accurate to order  $\epsilon$ . Consider again the eigenvalue problem for the Fokker-Planck equation:

$$\gamma(t)P(y, t) = D \frac{\partial^2 P(y, t)}{\partial y^2} + \frac{\partial}{\partial y} [U'(y, t)P(y, t)], \quad (27)$$

for  $y \in (a, \delta)$  (where  $a$  can be  $-\infty$  in some expressions below, but we will finally set  $a = -\delta$ ). We now consider  $t$  simply as a parameter in the eigenvalue problem (such that eigenvalue  $\gamma$  and eigenfunction  $P$  depend on the parameter  $t$  since the coefficient  $U'$  depends on  $t$ ). We will drop this parameter  $t$  throughout this subsection. The basic building block of solutions of (27) is the function  $\exp(-U(y)/D)$ , which we call

$$p(y) = \exp(-U(y)/D), \quad (28)$$

along with anti-derivatives of products of  $p$  of various orders, which we call

$$p_{s_1 \dots s_k}(y_0) = \int_{y_0}^{\delta} \dots \int_{y_{k-1}}^{\delta} \exp\left[\sum_{j=1}^k \frac{(-1)^{s_j} U(y_j)}{D}\right] dy_k \dots dy_1$$

(the subscripts  $s_j$  will be 1 or 2). We know that

$$\gamma = 0, \quad P_*(y) = p(y)p_2(y)/p_{12}(a) \quad (29)$$

solve (27) with the two boundary/integral conditions

$$P(\delta) = 0, \quad \int_a^{\delta} P(y) dy = 1. \quad (30)$$

The expression for  $P_*$  in (29), equals expression (18) for  $P_*$  with the specific normalization constant  $P_0 = J/D$  where  $J$  represents the probability flux. The probability flux  $J$  is the flow of probability per unit time per unit area.

The integral condition in (30) is based on an assumption that is only approximately correct if the noise level  $D$  is small: the probability flux  $J$  is constant in  $y$  such that the flux through the right boundary at  $+\delta$  must also enter at  $y = -\infty$  such that  $J$  is given by

$$J = \frac{D}{p_{12}(a)}. \quad (31)$$

The more appropriate boundary conditions for non-small noise level  $D$  result in the eigenvalue problem (27) for  $\gamma$  and  $P$  with

$$P(a) = 0, \quad P(\delta) = 0, \quad \int_a^{\delta} P(y) dy = 1,$$

which leads to a uniformly non-zero  $\gamma$ , including for the limit  $a \rightarrow -\infty$ .

We can express a first-order approximation of  $\gamma$  for non-small noise in terms of  $p$ , given in (28), by treating it as a perturbation of the small-noise limit and of  $\gamma = 0$ ,  $P = P_*$ . For a finite  $a \ll -1$  let us introduce the value of the solution  $P$  of (27) at  $a$  as a parameter  $\pi_a$ :

$$P(a) = \pi_a. \quad (32)$$

Then, we get a solution pair  $(\gamma, P)$  of (27) with boundary conditions (30), (32) for each small  $\pi_a$ . For  $\pi_a = P_*(a) = p(a)p_2(a)/p_{12}(a)$ , the solution is  $\gamma = 0$ ,  $P = P_*$ . So, to first order in  $\pi_a$ , we have for the parameter  $\pi_a = 0$

$$\gamma \approx -\gamma' P_*(a) = -\gamma' \frac{p(a)p_2(a)}{p_{12}(a)} \quad \text{with} \quad \gamma' = \left. \frac{d\gamma}{d\pi_a} \right|_{\pi_a = P_*(a)}. \quad (33)$$

The scalar  $\gamma'$  is part of the solution pair  $(\gamma', q)$  of the linearization of eigenvalue problem (27) with boundary conditions (30), (32), with respect to  $\pi_a$  in  $\pi_a = P_*(a)$ ,  $P = P_*$ ,  $\gamma = 0$ :

$$\gamma' P_* = \frac{\partial}{\partial y} \left( D \frac{\partial q}{\partial y} + U'(y)q \right) \quad (34)$$

with conditions

$$q(a) = 1, \quad q(\delta) = 0, \quad \int_a^{\delta} q(y) dy = 0.$$



This is an affine equation for  $\gamma'$  and  $q$ , which can be solved by integration, resulting in

$$\gamma' = \frac{Dp_{12}(a)^2}{p(a)[p_{12}(a)p_{212}(a) - p_2(a)p_{1212}(a)]},$$

such that the first-order estimate (33) gives

$$\gamma_1 := \gamma \approx \frac{D}{\frac{p_{1212}(a)}{p_{12}(a)} - \frac{p_{212}(a)}{p_2(a)}}. \quad (35)$$

Taking into account now that all quantities in (35) depend parametrically on time  $t$ , the probability  $\mathbb{P}_P$  of not following a path within a specific region is approximately

$$\mathbb{P}_P = 1 - \exp\left(-\int_{t_0}^{T_{\text{end}}} \gamma_1(t) dt\right), \quad (36)$$

where  $\gamma_1$  is given approximately in (35), when inserting  $-\delta$  for  $a$ . We can compare (36) with the simpler probability formula  $\mathbb{P}_J$  which is valid in the small noise limit

$$\mathbb{P}_J = 1 - \exp\left(-\int_{t_0}^{T_{\text{end}}} J(t) dt\right), \quad (37)$$

where the probability flux  $J$  is given in (31), when again inserting  $-\delta$  for  $a$ .

Section V will compare Monte-Carlo simulations, the numerical approximation using the first  $n$  instantaneous eigenmodes of the linear operator  $A(t)$  of the Fokker-Planck equation, the perturbation formula for the leading eigenvalue  $\gamma_1(t)$ , and the formula for the probability flux  $J = D/p_{12}(a)$  for the stationary density  $P_*$  in (29), which is accurate for small escape rates.

#### IV. SADDLE-NODE NORMAL FORM WITH PARAMETER RAMP AND NOISE

A prototypical model for rate-induced tipping was introduced by Ashwin *et al.* [5]. The time evolution of a scalar dependent variable  $x(t) \in \mathbb{R}$  is described by the saddle-node normal form equation:

$$\dot{x} = f(x, \lambda) = (x + \lambda)^2 - 1, \quad (38)$$

where w.l.o.g. we have set the normal form parameter to equal 1. The ODE (38) has two families of equilibria; one stable family  $W^{(s)}[0] = -\lambda - 1$  and one unstable family  $W^{(u)}[0] = -\lambda + 1$ . The parameter  $\lambda$  in equation (38) is assumed to be time dependent following a ramp given by:

$$\lambda(t) = \frac{\lambda_{\text{max}}}{2} \left( \tanh\left(\frac{\lambda_{\text{max}}\rho t}{2}\right) + 1 \right), \quad (39)$$

where  $\lambda_{\text{max}}$  determines how far the parameter  $\lambda$  is shifted and  $\rho$  adjusts the speed of the ramp. Equation (39)

can be described by an ODE for  $\lambda$  with the condition  $\lambda(0) = \lambda_{\text{max}}/2$ . Therefore the prototypical model can be described by the two dimensional ODE in the  $(x, \lambda)$  phase plane

$$\dot{x} = (x + \lambda)^2 - 1, \quad (40)$$

$$\dot{\lambda} = \rho\lambda(\lambda_{\text{max}} - \lambda). \quad (41)$$

For this system of ODEs a critical speed  $\rho = \rho_c = 4/[\lambda_{\text{max}}(\lambda_{\text{max}} - 2)]$  was found in [25], at which a heteroclinic connection  $(x, \lambda) = (-1 + (2/\lambda_{\text{max}} - 1)\lambda, \lambda)$  from  $(-1, 0)$  to  $(1 - \lambda_{\text{max}}, \lambda_{\text{max}})$  occurs (setting the critical rate  $\rho_c$  for rate-induced tipping). The time profile and phase portrait for  $\rho < \rho_c$  is presented in Figure 4. For a complete overview of all possible time profiles and phase portraits, see Ritchie and Sieber [10].

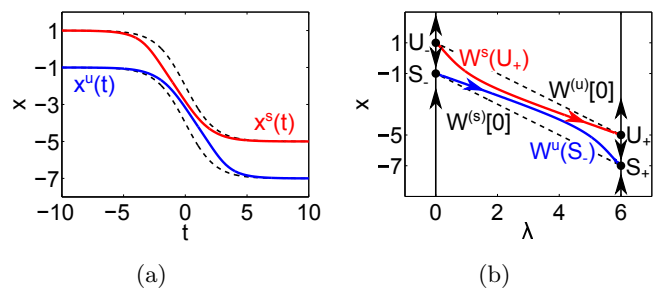


FIG. 4: Time profile (a) and phase plane (b) of system (40)–(41) for  $\rho = 0.14 < \rho_c$ . Black dashed curves are the stable ( $W^{(s)}[0] = -1 - \lambda$ ) and unstable ( $W^{(u)}[0] = 1 - \lambda$ ) branches of equilibria in the limit  $\rho = 0$ , blue and red curves are the unstable and stable manifolds,  $W^u(S_-)$  and  $W^s(U_+)$ , respectively ( $\lambda_{\text{max}} = 6$ ).

System (40)–(41) has 4 equilibria: two saddles  $S_- = (-1, 0)$ ,  $U_+ = (-5, 6)$ , one stable node  $S_+ = (-7, 6)$  and one unstable node  $U_- = (1, 0)$ ; see Figure 4b. The dashed lines  $W^{(s)}[0] = -1 - \lambda$  and  $W^{(u)}[0] = 1 - \lambda$  represent the family of stable and unstable equilibria for  $\rho = 0$  respectively. The curve  $W^u(S_-)$  is the unstable manifold of the saddle  $S_-$  and  $W^s(U_+)$  is the stable manifold of the saddle  $U_+$ . The time profile for  $x$  on the invariant manifolds  $W^u(S_-)$  and  $W^s(U_+)$ , denoted  $x^u(t)$  and  $x^s(t)$  respectively, is given in Figure 4a.

The manifold  $W^s(U_+)$  acts as a separatrix partitioning the plane into two distinct regions. Below  $W^s(U_+)$  all trajectories are attracted towards the stable node  $S_+$ , while any trajectories above the separatrix escape to  $+\infty$  in finite time.

We will continue using the parameters  $\rho$  and  $\lambda_{\text{max}}$  to remain consistent with the studies of [5, 10, 25]. The parameters  $\rho$  and  $\lambda_{\text{max}}$  (both are non-small) have the following relation to the parameters  $\epsilon$  and  $r$  of the sections II and III:

$$\epsilon = \frac{\rho\lambda_{\text{max}}}{4}, \quad r = \frac{\rho\lambda_{\text{max}}^2}{4}, \quad \lambda_{\text{max}} = \frac{r}{\epsilon}, \quad \rho = \frac{4\epsilon^2}{r}.$$

Using  $r$  and  $\epsilon$ , system (40)–(41) has the form

$$\dot{x} = (x + r\mu/\epsilon)^2 - 1, \quad \dot{\mu} = 4\epsilon\mu(1 - \mu),$$

the critical rate is  $r_c(\epsilon) = 1 + 2\epsilon$ , and the connecting orbit has the form  $(x, \mu) = (-1 + \mu(2 - r/\epsilon), \mu)$ . We will keep  $\lambda_{\max} = 6$  fixed and vary  $\rho$  between 0 and 1/6 which simultaneously varies  $\epsilon$  between 0 and 0.25 and  $r$  between 0 and 1.5. Therefore we will always have  $\epsilon \ll 1$  corresponding to gentle but long ramps.

The dynamics of  $x$  for the system (40)–(41), modified by adding noise to (40), are described by a stochastic differential equation (an example of the general equation (10)):

$$dX_t = [(X_t + \lambda(t))^2 - 1]dt + \sqrt{2D}dW_t. \quad (42)$$

The expressions (for example (36)) for probability of escape refer to a strip of half-width  $\delta$  around a deterministic reference trajectory  $\tilde{x}(t)$ . We choose the trajectory  $x^u(t)$  (blue curve in Figure 4a on the unstable manifold  $W^u(S_-)$ ) and a strip with a fixed width  $2\delta = 3$  around  $\tilde{x}(t)$ . Then we use the approximations for  $\mathbb{P}$ , (26), (36), (37) derived in section III to find the probability of escape.

As described in section III, we use co-moving coordinates

$$y(t) = x(t) - x^u(t),$$

such that the domain for  $y$ ,  $[-\delta, \delta]$  is fixed for all  $t$ . This transformation alters the ODE given in (40):

$$\begin{aligned} \dot{x} &= f(x, \lambda) = (x + \lambda)^2 - 1, & (43) \\ \dot{y} + \dot{x}^u &= f(y + x^u, \mu), \\ \dot{y} &= (y + x^u + \lambda)^2 - 1 - \dot{x}^u, \\ &= y^2 + \underbrace{2(x^u + \lambda)y}_{-c_1(t)} + \underbrace{(x^u + \lambda)^2 - 1 - \dot{x}^u}_{\text{equals 0}}, \end{aligned}$$

and so, instead of (40), we can express the new ODE as:

$$\dot{y}(t) = yg(y(t), t) = y^2(t) - c_1(t)y(t), \quad (44)$$

where  $c_1(t)$  is a time dependent scalar. Thus,  $yg(y, \mu(t))$  is used in the eigenvalue problem of the Fokker-Planck equation (27), with Dirichlet boundary conditions:

$$P(-\delta, t) = P(\delta, t) = 0.$$

Figure 5a gives the spectrum of the eigenvalues for different fixed times  $t$  and for noise level  $D = 0.06$ . The eigenvalues for the times when the system (40)–(41) is close to stationary (roughly all  $t \notin [-3, 3]$ ), is given in blue. The eigenvalues during the ramp (when  $\dot{\lambda}$  is of order 1) are given by the other colors, namely at  $t = 0$  (red

star),  $t = 1$  (black plus) and  $t = 2$  (green circle). All sets of eigenvalues  $\gamma_k$  are on parabolas ( $\gamma_k = \text{Re}\gamma_k \sim -k^2$ ) (Figure 5a), with a nearly linear relationship  $\gamma_k \sim -k$  for the dominant eigenvalues (Figure 5b) and  $\gamma_1(t) \approx 0$  for all  $t$ .

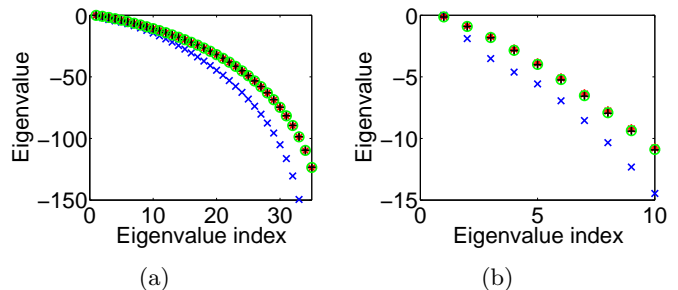


FIG. 5: Eigenvalue spectrum of the linear operator  $A(t)$  for the Fokker-Planck equation corresponding to the system (40)–(41), with  $\rho = 0.14$  ( $\epsilon = 0.21$ ,  $r = 1.26$ ),  $D = 0.06$  and at times  $t = -10$  (blue cross),  $t = 0$  (red star),  $t = 1$  (black plus) and  $t = 2$  (green circle).

Notice, that the first three eigenvalues during the ramp are all greater than the second eigenvalue for times when the system is close to stationary, for example at  $t = -10$  (blue). Thus, the contribution of additional modes is more significant during the period of the shift. The computational study in Section V will compare single- and three-mode approximations.

We return to Figure 3 which displays different types of densities to be compared with the reference probability density from simulations in blue. We will focus on the comparison between the single- (green (light gray)) and three-mode approximations (red (dark gray)) to the reference probability density for ramping speed  $\rho = 0.14$  ( $\epsilon = 0.21$ ,  $r = 1.26$ ). According to our analysis all differences between approximations and simulation results are caused by the non-zero  $\epsilon$ .

Initially both the single-mode and three-mode approximation match the probability density well (not shown), however, a visible deviation appears in the single-mode at  $t = 0$ , see panel (b). The single-mode approximation develops a larger tail than the density from simulations (and the three-mode approximation). This corresponds to an overestimation of the escape and hence, the peak of the density is lower. The density has also shifted further along the  $x$ -axis because the density instantaneously adjusts to the effective potential, as previously discussed. In contrast the three-mode approximation is still providing a good match to the reference probability density. For larger times the single-mode is underestimating the escape compared to the simulations. The single-mode approximation converges back to the reference density for  $t > 2$  (panel (d)) when the shift slows down. The three-mode approximation follows the reference density with greater accuracy throughout.



## V. SYSTEMATIC PARAMETER STUDY IN $\rho$ AND $D$

In this section, we will compare different approximations for the probability of escape (noise and rate-induced tipping) in the two parameter  $(\rho, D)$  - plane. We will use Monte-Carlo simulations (described below) as the reference for the probability of escape. Section III proposed two approximate expressions for the single-mode approximation:

- the escape probability based on the probability flux  $J = D/p_{12}(a)$  with  $a = -\delta$ , given in (37);
- the escape probability based on the first-order approximation of the leading eigenvalue  $\gamma_1$ , given in (36) (also with  $a = -\delta$ ).

We also compute the escape probabilities based on single- and three-mode approximations, by solving the ODE eigenvalue problem (27) with Dirichlet boundary conditions at  $-\delta$  and  $\delta$  numerically for the first  $n$  modes  $((\gamma_k, v_k), k = 1, \dots, n)$  together with the ODE (24), such that the escape probability is given by (26) for  $n = 1$  and  $n = 3$ .

Figure 6a shows the probability of noise and rate-induced tipping occurring in the two parameter  $(\rho, D)$  - plane, calculated using Monte-Carlo simulations. This has been performed by starting with a large number of realizations at  $x_0 = -1$  at  $t_0 = -10$  and evolving according to the SDE (10). The fraction of realizations that pass  $x_T = 4$ , and, hence, go to  $+\infty$  in finite time, approximates the probability of tipping (or probability of escape). The reference probability is not derived from the number of realizations escaping the strip  $\{y(t) \in [-\delta, \delta] : t \in [t_0, T_{\text{end}}]\}$  but by the fraction of realizations that have crossed an arbitrary line  $x_T = 4$  (the choice of  $\delta$  and  $x_T$  is such that this difference has a small effect).

Figure 6a shows the probability of escape (in %) for all ramping speeds  $\rho$  up to  $\rho_c = 1/6$  and a range of noise levels  $D$ . The color contours indicate that the probability of escape is small for small  $\rho$  and  $D$ . As  $\rho$  increases towards  $\rho_c$  and the noise level increases so does the probability of escape, reaching approximately 70% probability of escape for  $\rho = \rho_c$  and  $D = 0.2$ .

*a. Region in  $(\rho, D)$ -plane considered* We can expect the single-mode (or three-mode) approximations to be accurate only in a range of parameters  $\rho$  up to a value  $\rho_{\text{max}} = 0.14$  that is slightly smaller than the critical value  $\rho_c = 1/6$  (where tipping occurs without noise). The reason for this is in the error terms when replacing the dynamic Fokker-Planck equation for the density with its projection onto leading time-dependent mode(s). These error terms are only small if the time derivative of the (time-dependent) drift  $yg(y, t)$  in (20) is small.

As introduced in (19) in Section III A, the eigenvalue problem for the Fokker-Planck equation is solved in a co-moving coordinate system along the path  $\tilde{x}(t)$  ( $y =$

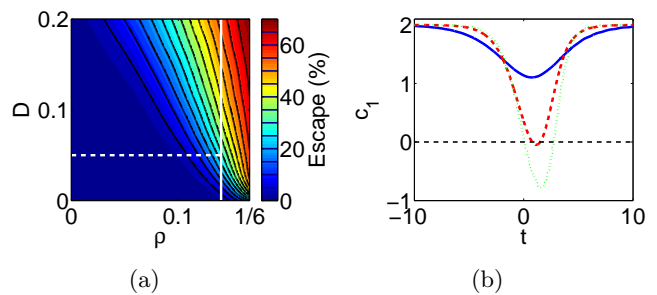


FIG. 6: (a) Overview probability of escape in the  $(\rho, D)$  - plane using simulations (with smoothing), where contours are spaced for every 5% of escape. Vertical white line indicates value of  $\rho$  such that  $\min(c_1) = 0$  and horizontal dashed line shows lowest value of  $D$  for which the probability can be calculated for the modes. (b) Time profile of the critical points of  $U(y, t)$ , one of them is always  $y = 0$ , the other is  $c_1(t)$  for  $\rho = 0.08$  (blue solid), 0.14 (red dashed) and 0.16 (green dotted).

$x - \tilde{x}(t)$ ) such that the path is centered at  $y = 0$  within the fixed domain  $y \in [-\delta, \delta]$  and the drift is given by  $yg(y, t) = y(y - c_1(t))$  (where  $c_1(t)$  is given in (43)). Its potential  $U(y, t) = -\int y(y - c_1(t))dy$  has a well at  $y = 0$  and a hill top at  $c_1$  for  $c_1 > 0$ , but a hill top at  $y = 0$  (and a well at  $c_1$ ) for  $c_1 < 0$ . Figure 6b illustrates the time profile of  $c_1(t)$  for different values of the drift speed  $\rho$ . The limit of  $c_1(t)$  for  $t \rightarrow \pm\infty$  is 2 such that  $c_1(t) \approx 2$  for  $t$  close to  $t_0 = -10$  and  $T_{\text{end}} = 10$ . For small drift speeds  $\rho$  the deviations of  $c_1$  from its asymptotic value are small (blue solid curve in Figure 6b), while for  $\rho = 0.16$   $c_1(t)$  becomes negative for some time interval, making the trajectory  $x(t) = \tilde{x}(t)$  or  $y(t) = 0$  locally repelling (green dotted curve in Figure 6b).

The error of the single-mode approximation is small if  $|\dot{c}_1(t)|$  is small, which is the case for  $t$  near  $t_0$  and  $T_{\text{end}}$ , and for  $t$  close to the minimum of  $c_1$ . If  $c_1(t) > 0$  for all  $t$  then the time  $t_{\text{min}}$ , when  $c_1(t)$  is minimal, correspond to those times where escape is most likely to occur, since at these times the potential barrier is smallest. At times near  $t_{\text{min}}$  the mode approximation error is also small since  $|\dot{c}_1(t)|$  is small.

However, if  $c_1(t) < 0$  for a range of  $t$  then escape occurs with a non-small probability at times when  $|\dot{c}_1(t)|$  is not small, leading to an error in the single-mode approximation that is not small.

Hence, we choose a range for the parameter  $\rho$  such that  $c_1(t)$  stays positive along the entire path for all  $\rho$ .

We also remove small values of  $D$  ( $D < 0.05$ ) from our consideration, since for  $\rho < \rho_c$ , but not close to  $\rho_c$ , escape probabilities are small compared to errors in Monte-Carlo simulations and in the numerical computations of the integrals needed for  $\gamma_1$  in (35). In this region the probability of escape is exponentially small in  $D$  (that is, of order  $\exp(-C/D)$  for some constant  $C > 0$ ).

Figure 7a shows the reference probability - the proba-

bility of escape calculated using Monte-Carlo simulations for this restricted region (with a slightly different color scale to Figure 6a). The remaining three panels of Figure 7 give the signed error of the approximation, compared to the reference, in percent. In the color scale for these panels, a green (light gray) color represents good agreement between the approximation and reference escape. A positive error (red) means an overestimation and a negative error (blue) corresponds to an underestimation when using the approximation method.

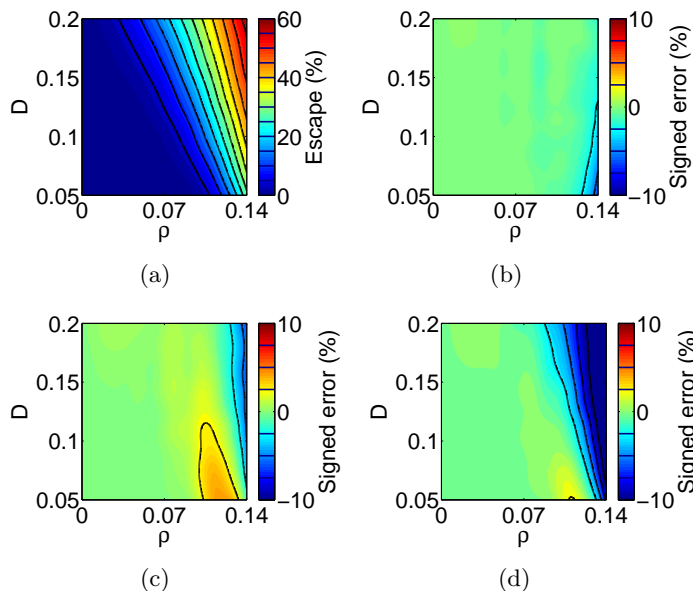


FIG. 7: (a) Reference probability of escape in  $(\rho, D)$  - parameter plane, obtained using Monte-Carlo simulations (with smoothing) - observing the fraction of realizations that escape the potential landscape. (b)-(d) Evaluating approximation methods for probability of escape by plotting in color the signed error % between the approximation and the reference escape. Positive error (red) reflects an overestimation and negative (blue) an underestimation. Approximation methods used: (b) Numerically calculate first three ( $n = 3$ ) modes of the linear operator of the Fokker-Planck equation (26). (c) Analytical single-mode approximation (36) and (d) analytical probability flux approximation (37). Contours are spaced at 5% intervals for panel (a) and 2.5% intervals for panels (b)-(d) with the zero contour omitted.

Figure 7b shows the probability of escape calculated using (26) with the first  $n = 3$  instantaneous eigenmodes of the linear operator of the Fokker-Planck equation. The three-mode approximation was computed by solving the ODE eigenvalue problem (22) numerically. It approximates the reference probability of escape over the specified region well, except for  $\rho$  close to 0.14 and small noise levels.

Figures 7c and 7d compare the single-mode approxi-

mation (36) and the approximation using the probability flux (37) to the reference escape. The single-mode approximation offers a very good agreement with the reference escape for  $\rho < 0.1$ . For larger  $\rho$  the formula gives an overestimation of the escape for small noise levels. The probability flux again approximates the probability for small  $\rho$  values well, but when the probability of escape increases to above 20% the probability flux underestimates the reference escape.

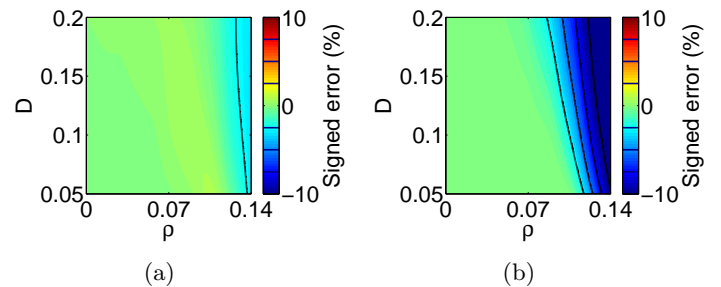


FIG. 8: Color plots of the signed error between the probability calculated numerically from the ( $n = 1$ ) single-mode (26) and (a) the single-mode approximation (36) or (b) the probability flux approximation (37). A positive error (red) corresponds to the prescribed approximation overestimating the numerical probability calculated from the single-mode, whereas a negative error (blue) represents an underestimation. Contours are spaced at 2.5% intervals with the zero contour omitted.

Figure 8a compares the single-mode approximation formula (36) to the exact single-mode approximation, solving (22) and (26) for  $n = 1$  numerically. We see that the difference is much smaller than the error caused by the approximation of the Fokker-Planck equation with a single mode, see Figure 7c.

In contrast, the probability flux (37) systematically underestimates the escape probability when it is greater than 20%; see Figure 8b. This is as expected because the estimate (37) based on a spatially constant probability flux assumes the flux escaping at the boundary  $+\delta$  re-enters at the boundary  $-\delta$ . The difference between the two estimates is larger when the escape probability is high, because the assumption of spatially constant flux is only approximately true if escape is sufficiently rare compared to the time it takes for realizations to reach the potential well from the boundary at  $-\delta$ . This leads to an underestimate for those values of  $D$  and  $\rho$  when the probability of escape is greatest.

## VI. DISCUSSION

We have provided approximations for the critical rate for deterministic rate-induced tipping and for the probability of noise-induced tipping during parameter shifts.

These approximations are valid for parameter shifts that are asymptotic to constant parameter values for  $t \rightarrow \pm\infty$  and that are “long but gentle”: the small parameter  $\epsilon$  is the ratio between maximal ramp (shift) speed  $r$  and the length of the parameter shift. The deterministic critical rate  $r_c(\epsilon) = r_0 + \epsilon r_1 + O(\epsilon^2)$  is an order- $\epsilon$  perturbation from the critical rate  $r_0$  for constant-speed parameter shifts (which were discussed in [5]). The approximation for the tipping probability in the presence of noise is based on the instantaneous eigenmode expansion of the linear operator for the Fokker-Planck equation to approximate the quasi-stationary probability densities in a strip  $S_\delta$  of half-width  $\delta$  around the deterministic trajectory (a connecting orbit between equilibria). Moreover, we have derived a general perturbation formula to calculate the leading eigenvalue, which approximates the probability calculated from a single eigenvalue (the single-mode approximation) and thus, gives a good approximation to the reference probability of tipping for small  $\epsilon$ .

The limitation of the proposed estimate using single-mode approximation is that it fails for some maximal ramp speeds  $r$  less than  $r_c$ , even for small noise levels  $D$ . A brief derivation in Appendix A shows that the single-mode approximation is generally valid for parameters up to  $r = r_0 + \epsilon r_1/2$  such that the escape rate from the strip  $S_\delta$  is maximal during times when the time-derivative of the shape of the underlying potential well is minimal.

## ACKNOWLEDGMENTS

P.D.L.R.’s research was supported by funding from the EPSRC Grant EP/M008495/1, J.S. research was supported by funding from the EPSRC Grant EP/N023544/1 and by funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement number 643073.

### Appendix A: Expansion of the critical rate in the small parameter

This section presents in more detail the expansion of the critical rate for systems with a ramped parameter shift. The critical rate is defined as the threshold at which a system fails to track the continuously changing quasi-steady state and thus generating rate-induced tipping.

The simplest example of rate-induced tipping is a system subjected to a linear parameter shift  $r_{\text{lin}}$ ,  $\dot{x} = f(x + b\lambda)$  with  $\lambda = r_{\text{lin}}t$ . We assume that the linearly shifted system ( $y = x + b\lambda$ )

$$\dot{y} = f(y) + r_{\text{lin}}b \quad (\text{A1})$$

( $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ ,  $y(t) \in \mathbb{R}^n$ ) has a generic saddle-node bifurcation at  $r_{\text{lin}} = r_0 > 0$ ,  $y = y_0$ . We assume that the number of unstable dimensions of the equilibria changes from 0 to 1 at the saddle-node and that the equilibria

exist for  $r_{\text{lin}} < r_0$ . In this section we derive the first-order expansion of the critical rate  $r_c$  in  $\epsilon$  for the system

$$\dot{y} = f(y) + b(r\Gamma(\mu) - r_0), \quad \dot{\mu} = \epsilon\Gamma(\mu). \quad (\text{A2})$$

System (A2) describes the scenario of a ramped shift  $\dot{x} = f(x + br\mu/\epsilon)$  with maximal speed  $r$ , again in shifted coordinates  $y = x + br\mu/\epsilon$ . The direction of the shift is determined by  $b \in \mathbb{R}^n$  and  $\Gamma(\mu)$  satisfies the following properties

$$\begin{aligned} 0 &= \Gamma(0) = \Gamma(1), \\ 1 &= \Gamma(\mu_{\text{crit}}) = \max\{\Gamma(\mu) : \mu \in [0, 1]\}, \text{ (this defines } \mu_{\text{crit}}\text{),} \\ 0 &> \Gamma''(\mu_{\text{crit}}), \text{ (let } g_2 := -\Gamma''(\mu_{\text{crit}})/2\text{),} \\ \Gamma(\mu) &\in (0, 1) \text{ for all } \mu \in (0, \mu_{\text{crit}}) \text{ and } \mu \in (\mu_{\text{crit}}, 1) \end{aligned} \quad (\text{A3})$$

such that  $\Gamma$  has a unique non-degenerate maximum at  $\mu_{\text{crit}}$ . We make a change of coordinates  $y = y_0 + v_0z$  to shift the origin in (A2) to the saddle-node bifurcation at  $y = y_0$ :

$$v_0\dot{z} = f(y_0 + v_0z) + b(r\Gamma(\mu) - r_0), \quad (\text{A4})$$

where  $v_0$  is the right nullvector of  $\partial f(y_0)$  (which has a one-dimensional nullspace by the assumption of a saddle-node bifurcation at  $y = y_0$  for (A1)). Furthermore, this assumption implies

$$f(y_0 + v_0z) = \frac{1}{2}z^2\partial^2 f(y_0)v_0^2 + \mathcal{O}(z^3), \quad (\text{A5})$$

Inserting (A5) into (A4) and applying  $w_0^T$  (the left nullvector of  $\partial f(y_0)$ , scaled such that  $w_0^T v_0 = 1$ ) to (A4) gives

$$\dot{z} = a_0(r\Gamma(\mu) - r_0) + a_2z^2 + \mathcal{O}(z^3),$$

where  $a_0 = w_0^T b$ ,  $a_2 = \frac{1}{2}w_0^T \partial^2 f(y_0)v_0^2$  are both non-zero by the assumption of a generic saddle-node. As we assume that the equilibria for fixed  $\mu$  exist for  $r\Gamma(\mu) < r_0$ , we can choose the orientation of  $v_0$  such that  $a_0 > 0$  and  $a_2 > 0$ . Thus, the reduced autonomous system (A2) has the form (for small  $z \in \mathbb{R}$ )

$$\dot{z} = a_0(r\Gamma(\mu) - r_0) + a_2z^2 + \mathcal{O}(z^3), \quad (\text{A6})$$

$$\dot{\mu} = \epsilon\Gamma(\mu). \quad (\text{A7})$$

We zoom into the neighborhood of the maximal rate of change of  $\Gamma(\mu)$  (at  $\mu_{\text{crit}}$ ),  $z = 0$  and  $r = r_0$  by introducing rescaled variables and time

$$\begin{aligned} \mu_{\text{old}} &= \mu_{\text{crit}} + \sqrt{\epsilon c_m} \mu_{\text{new}}, & \text{where } c_m &= [g_2 r_0 a_0 a_2]^{-1/2}, \\ z_{\text{old}} &= \sqrt{\epsilon c_z} z_{\text{new}}, & \text{where } c_z &= [g_2 r_0 a_0 / a_2^3]^{1/2}, \\ r_{\text{old}} &= r_0 + \epsilon c_r r_{\text{new}}, & \text{where } c_r &= [g_2 r_0 / (a_0 a_2)]^{1/2}, \\ t_{\text{old}} &= \sqrt{c_t / \epsilon} t_{\text{new}}, & \text{where } c_t &= [g_2 r_0 a_0 a_2]^{-1/2}, \end{aligned}$$

and expanding  $\Gamma(\mu)$  near its unique maximum in  $\mu_{\text{crit}}$  (recall from (A3) that  $\Gamma(\mu_{\text{crit}}) = 1$  and  $g_2$  is defined as  $-\Gamma''(\mu_{\text{crit}})/2$ ):

$$\Gamma(\mu_{\text{old}}) = \Gamma(\mu_{\text{crit}} + \sqrt{\epsilon c_m} \mu_{\text{new}}) = 1 - \epsilon g_2 c_m \mu_{\text{new}}^2 + o(\epsilon).$$

In these coordinates the extended system (A6)–(A7) can then be written as

$$\dot{z} = z^2 + r - \mu^2 + o(1), \quad (\text{A8})$$

$$\dot{\mu} = 1 + O(\epsilon) \quad (\text{A9})$$

Orbits that stay close to the family of equilibria of (A1) uniformly for all  $\epsilon \rightarrow 0$  are perturbations of orbits that exist for all time in the limiting system of (A8)–(A9) for  $\epsilon = 0$ . We have 3 cases for (A8)–(A9) with  $\epsilon = 0$ . In all 3 cases there exists a unique *globally defined* orbit  $z(\mu)$  that exists for all times.

$r < 1$  (Tracking): the globally defined orbit  $z(\mu)$  has the limiting behavior  $z(\mu) + |\mu| \rightarrow 0$  for  $\mu \rightarrow \pm\infty$  and is stable forward in time. All orbits starting with  $z < 0$  and  $\mu \ll -1$  converge to the globally defined orbit.

$r = 1$  (Critical): the globally defined orbit is  $z(\mu) = \mu$ .

$r > 1$  (Escape): the globally defined orbit  $z(\mu)$  has the limiting behavior  $z(\mu) - |\mu| \rightarrow 0$  for  $\mu \rightarrow \pm\infty$  and is stable backward in time. All orbits starting with  $z < 0$  and  $\mu \ll -1$  diverge to  $+\infty$  in finite time after  $\mu > -\sqrt{r}$ .

In the original coordinates the rescaled parameter  $r = 1$  equals the first order expansion for the critical rate  $r_c(\epsilon)$

$$r_c(\epsilon) = r_0 + \epsilon \sqrt{-\frac{r_0 \Gamma''(\mu_{\text{crit}})}{2a_0 a_2}}.$$

If white noise of variance  $\sigma^2$  is added to (A6):

$$dz = [a_0(r\Gamma(\mu) - r_0) + a_2 z^2 + O(z^3)]dt + \sigma dW_t,$$

then  $\sigma^2$  needs to be of the scale  $\epsilon^{3/2}$ . If  $\sigma^2 = 2D[\epsilon c_n]^{3/2}$  with  $c_n = (g_2 r_0 a_0)^{1/2} a_2^{-5/6}$  then the rescaled equation for  $z$  is

$$dz = [z^2 + r - \mu^2 + o(1)]dt + \sqrt{2D}dW_t.$$

*Limitation of single-mode approximation* We use the saddle-node normal form (A8)–(A9) to provide insight into the limitation of the single-mode approximation in reference to the maximal ramp speed  $r$ . The unique globally defined orbit  $\tilde{z}(t)$  of (A8)–(A9) for  $\epsilon = 0$  is used to change to a co-moving coordinate system

$$y(t) = z(t) - \tilde{z}(t),$$

with respect to  $y$  such that (A8)–(A9) has the form

$$\begin{aligned} \dot{z} &= z^2 + r - t^2, \\ \dot{\tilde{z}} + \dot{y} &= (\tilde{z} + y)^2 + r - t^2, \\ \dot{y} &= \tilde{z}^2 + 2\tilde{z}y + y^2 + r - t^2 - \dot{\tilde{z}}, \\ \dot{y} &= y^2 + 2\tilde{z}y = y(y + 2\tilde{z}). \end{aligned} \quad (\text{A10})$$

The potential  $U(y, t) = -\int y(y + 2\tilde{z}) dy$  in the new coordinate system has a well at  $y = 0$  and a hill top at  $y = -2\tilde{z}$  for  $\tilde{z} < 0$ , but a hill top at  $y = 0$  (and a well at  $y = -2\tilde{z}$ ) for  $\tilde{z} > 0$ .

If white noise is added to (A10), the error of the single-mode approximation is small provided  $|\dot{\tilde{z}}(t)|$  is small, which is the case for  $t$  close to the maximum of  $\tilde{z}$ . If  $\tilde{z}(t) < 0$  for all  $t$  then escape is most likely to occur close to time  $t_{\text{max}}$ , the maximum of  $\tilde{z}(t)$  since the potential barrier is at its lowest. The mode approximation error at times close to  $t_{\text{max}}$  is small because  $|\dot{\tilde{z}}(t)|$  is small, as discussed in the paper.

However, if  $\tilde{z}(t) > 0$  for a range of  $t$  then escape occurs with a non-small probability at times when  $|\dot{\tilde{z}}(t)|$  is not small. We identify that a maximal ramp speed  $r \approx 0.59$  corresponds to  $\max\{\tilde{z}(t) : t \in [t_0, T_{\text{end}}]\} = 0$ . Therefore, in the original coordinates the single-mode approximation fails for some maximal ramp speeds  $r$  less than  $r_c$ , and in particular, we consider only for parameters up to

$$r = r_0 + \frac{1}{2}\epsilon c_r,$$

where in the paper the constant  $r_1$  is the same as  $c_r$ .

## Appendix B: Phase planes of shifted system

In this section we consider all qualitatively different phase planes for the shifted slow-fast autonomous system

$$\dot{y} = f(y) + br\Gamma(\mu), \quad (\text{B1})$$

$$\dot{\mu} = \epsilon\Gamma(\mu). \quad (\text{B2})$$

Again we assume that in the limit  $\epsilon = 0$ , (B1) has a saddle-node bifurcation at  $r\Gamma(\mu) = r_0 > 0$ ,  $y = y_0$  with a stable branch  $y^{(s)}[r\Gamma(\mu)]$  and an unstable branch  $y^{(u)}[r\Gamma(\mu)]$  of equilibria emerging for  $r\Gamma(\mu) \in [0, r_0]$ . The properties (A3) of  $\Gamma$  imply that there exists three distinct sets of equilibria branches depending on the value of  $r$  in relation to  $r_0$ . These are presented in the  $(\mu, y)$  - phase plane in panel (a) for  $r < r_0$ , panel (b) for  $r = r_0$  and panels (c)–(e) for  $r > r_0$  in Figure 9. We will discuss the differences between panels (c)–(e) but initially we will just focus on the branches of stable (blue dashed) and unstable (red dashed) equilibria for  $\epsilon = 0$ .

For  $r < r_0$  in the limit  $\epsilon = 0$  there exists one stable equilibrium and one unstable equilibrium for all  $\mu \in [0, 1]$ , see panel (a). This means there is a continuous branch of stable equilibria connecting  $(y, \mu) = (y^{(s)}[0], 0)$  and  $(y^{(s)}[0], 1)$  and likewise an unstable branch connecting  $(y, \mu) = (y^{(u)}[0], 0)$  and  $(y^{(u)}[0], 1)$ . The phase portrait in panel (a) corresponds to the case  $r < 0$  for the rescaled  $r$  in the rescaled system (A8)–(A9).

For  $r = r_0$  the stable and unstable equilibria meet at  $\mu_{\text{crit}} \in (0, 1)$  where  $\Gamma(\mu_{\text{crit}}) = 1$ ; see panel (b). In contrast to panel (a) the branches of equilibria approach at a linear rate and then also move away at a linear rate. This

therefore means that the branches are non-differentiable at  $\mu = \mu_{\text{crit}}$  but the continuous connections still exist.

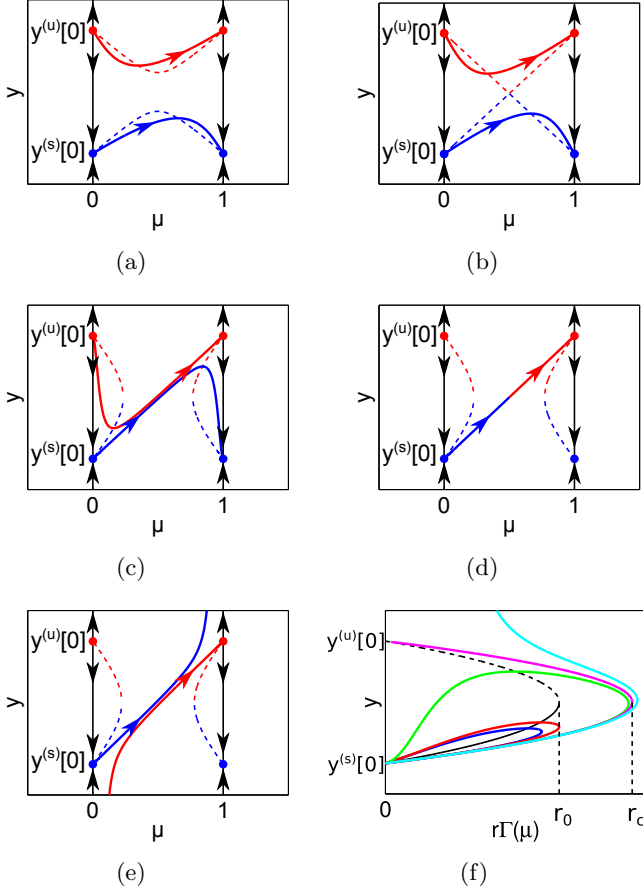


FIG. 9: (a)–(e) Qualitatively different phase planes of system (B1)–(B2) dependent on the value of the maximal ramp speed  $r$  in relation to the saddle-node bifurcation at  $r\Gamma(\mu) = r_0$  and the critical rate  $r_c$  that induces tipping (see panel (f)). (a)  $r < r_0 < r_c$ , (b)  $r = r_0 < r_c$ , (c)  $r_0 < r < r_c$ , (d)  $r_0 < r = r_c$ , (e)  $r_0 < r_c < r$ . Blue (lower) and red (upper) dashed lines are the stable and unstable branches of equilibria in the limit  $\epsilon = 0$ . Solid blue and red curves represent the unstable manifold of the saddle  $(y, \mu) = (y^{(s)}[0], 0)$  and the stable manifold of the saddle  $(y, \mu) = (y^{(u)}[0], 1)$  respectively. Panel (f) provides the bifurcation diagram of (B1)–(B2) in the  $(r\Gamma(\mu), y)$  - plane. Black solid curve denotes branch of stable equilibria and black dashed curve branch of unstable equilibria. Superimposed on top are trajectories of a realisation starting arbitrarily close to  $(y, \mu) = (y^{(s)}[0], 0)$  for panel (a) given in blue, (b) red, (c) green, (d) pink, (e) light blue (listed in increasing values of  $r$ ). Parameter:  $b = 1$ .

In panels (c)–(e) the maximal ramp speed  $r$  is greater than  $r_0$  meaning that for  $\mu$  in some interval  $(\mu_l, \mu_u) \subset (0, 1)$  centered around  $\mu_{\text{crit}}$  no equilibria exist in the limit  $\epsilon = 0$ . Instead two saddle-node bifurcations form at  $\mu_l$

and  $\mu_u$  with the branches of stable and unstable equilibria emerging for  $\mu \in [0, \mu_l]$  and  $\mu \in [\mu_u, 1]$  to  $y^{(s)}[0]$  and  $y^{(u)}[0]$  respectively at  $\mu = 0$  and 1.

Let us now discuss the solid blue and red curves in panels (a)–(e) which denote the unstable manifold of  $(y, \mu) = (y^{(s)}[0], 0)$  and the stable manifold of  $(y, \mu) = (y^{(u)}[0], 1)$  respectively. Panels (a)–(c) depict the tracking scenario ( $r < r_c$ ) such that a connecting orbit exists between  $y^{(s)}[0]$  at  $\mu = 0$  and  $y^{(s)}[0]$  at  $\mu = 1$ , the unstable manifold of  $(y, \mu) = (y^{(s)}[0], 0)$ . The stable manifold of  $(y, \mu) = (y^{(u)}[0], 1)$  acts as a separatrix where all solutions below the manifold converge to the stable node  $(y, \mu) = (y^{(s)}[0], 1)$  and all those above escape to infinity. For panels (a) and (b), solutions  $(y(t), \mu(t))$  starting close to  $(y^{(s)}[0], 1)$  stay close to the stable equilibrium branch  $y^{(s)}[r\Gamma(\mu(t))]$  for all  $t$  (distance goes to 0 as  $\epsilon \rightarrow 0$ ). Notice in panel (c) that although  $r > r_0$  such that the saddle-node bifurcation is crossed for a small period of time the system (B1)–(B2) still does not tip because  $r < r_c = r_0 + \mathcal{O}(\epsilon)$  the critical rate. The scenario in panel (c) corresponds to the case  $0 < r < 1$  for the rescaled  $r$  in the rescaled system (A8)–(A9).

In panel (d)  $r$  equals  $r_c$ , which creates a saddle-to-saddle connection from  $(y^{(s)}[0], 0)$  to  $(y^{(u)}[0], 1)$ . Finally, in panel (e)  $r > r_c$  which induces system (B1)–(B2) to tip. This means that initial conditions starting arbitrarily close to  $(y^{(s)}[0], 0)$  go on to escape following the unstable manifold of  $(y^{(u)}[0], 1)$ . Panels (d) and (e) correspond to the critical case  $r = 1$  (d) and the escape case  $r > 1$  (e) for the rescaled  $r$  in the rescaled system (A8)–(A9).

We present in panel (f) the bifurcation diagram of (B1)–(B2) in the  $(r\Gamma(\mu), y)$  - plane. The black solid and dashed curves give the stable and unstable branches of equilibria respectively. Superimposed on top are colored curves representing the trajectory for starting close to  $(y, \mu) = (y^{(s)}[0], 0)$  for each of the scenarios in panels (a)–(e). The structure of the ramp means that solutions starting at  $(r\Gamma(\mu), y) = (0, y^{(s)}[0])$  approach the saddle-node at  $r\Gamma(\mu) = r_0$  at a slow speed initially which gets faster until  $\mu = \mu_{\text{crit}}$  is reached where  $r\Gamma(\mu_{\text{crit}}) = r$  (the turning points of the colored curves; recall,  $\mu_{\text{crit}}$  is where  $\max\{\Gamma(\mu) : \mu \in [0, 1]\} = 1$ ). The solutions then return back to  $r\Gamma(1) = 0$ .

A prominent feature in all cases is that initially the trajectory appears to lag behind the stable equilibrium branch  $y^{(s)}[r\Gamma(\mu)]$  as  $\Gamma(\mu(t))$  increases. The dark blue trajectory corresponds to panel (a) where  $r < r_0$  and so in this scenario the saddle-node is not reached, which means the stable quasi-steady state is always present. Not evident from the bifurcation diagram is that when  $\Gamma(\mu(t))$  reaches its maximum value the shift is at its fastest. Hence, the trajectory crosses the branch  $y^{(s)}[r\Gamma(\mu(t))]$  since the quasi-steady state changes direction quickly in comparison to the trajectory. The trajectory continues to lag until the ramp comes to a rest and the trajectory returns to  $y^{(s)}[0]$  at  $r\Gamma(1) = 0$ . A similar pattern is observed for the red trajectory where  $r = r_0$  (the scenario from panel (b)). The lag becomes more pro-

nounced the closer  $r\Gamma(\mu(t))$  gets to  $r_0$  the saddle-node. Thus, only touching the saddle-node and not crossing it, combined with the speed of the shift, the trajectory will not escape and instead again cross the stable branch  $y^{(s)}[r\Gamma(\mu(t))]$  and converge back to  $y^{(s)}[0]$  at  $r\Gamma(1) = 0$ .

The green curve shows the trajectory for  $r_0 < r < r_c$ , which corresponds to panel (c). For  $r\Gamma(\mu(t)) > r_0$  the trajectory gives the impression that it begins to escape but only slowly. Though once again this is deceptive since the speed of the ramp is at its fastest during this phase and so it is only a short period of time before  $\Gamma(\mu(t)) = r_0$  for a second time. The quasi-steady states then begin to emerge but the trajectory is now above

the unstable quasi-steady state. The trajectory continues to escape until it crosses the branch  $y^{(u)}[r\Gamma(\mu(t))]$ . This implies that the quasi-steady state is shifting faster than the trajectory is escaping. Once across the unstable quasi-steady state the trajectory gets attracted back to  $y^{(s)}[0]$  at  $r\Gamma(1) = 0$  to complete the connecting orbit.

Whereas, the pink trajectory ( $r = r_c$ , panel (d)) is shifted a little bit further such that the trajectory meets the  $y^{(u)}[r\Gamma(\mu(t))]$  branch only at the end of the ramp. The last scenario,  $r > r_c$ , the light blue trajectory is shifted sufficiently past the saddle-node bifurcation, such that the trajectory has enough time to escape before the system recovers.

- 
- [1] F. Kwasniok, *Physical Review E* **88**, 052917 (2013).
- [2] C. Folke, S. Carpenter, B. Walker, M. Scheffer, T. Elmqvist, L. Gunderson, and C. Holling, *Annual Review of Ecology, Evolution, and Systematics*, 557 (2004).
- [3] M. Hirota, M. Holmgren, E. H. Van Nes, and M. Scheffer, *Science* **334**, 232 (2011).
- [4] R. M. May, S. A. Levin, and G. Sugihara, *Nature* **451**, 893 (2008).
- [5] P. Ashwin, S. Wiczorek, R. Vitolo, and P. Cox, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **370**, 1166 (2012), Correction co-authored with C. Perryman (Née Hobbs) **371**, 20130098 (2013).
- [6] T. M. Lenton, *Annual Review of Environment and Resources* **38**, 1 (2013).
- [7] P. Ashwin, C. Perryman, and S. Wiczorek, arXiv preprint arXiv:1506.07734 (2015).
- [8] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. Van Nes, M. Rietkerk, and G. Sugihara, *Nature* **461**, 53 (2009).
- [9] I. A. van de Leemput, M. Wichers, A. O. Cramer, D. Borsboom, F. Tuerlinckx, P. Kuppens, E. H. van Nes, W. Viechtbauer, E. J. Giltay, S. H. Aggen, *et al.*, *Proceedings of the National Academy of Sciences* **111**, 87 (2014).
- [10] P. Ritchie and J. Sieber, *Chaos* **26**, 093116 (2016), <http://dx.doi.org/10.1063/1.4963012>.
- [11] M. Scheffer, S. R. Carpenter, T. M. Lenton, J. Bascompte, W. Brock, V. Dakos, J. Van De Koppel, I. A. Van De Leemput, S. A. Levin, E. H. Van Nes, *et al.*, *science* **338**, 344 (2012).
- [12] T. Lenton, V. Livina, V. Dakos, E. Van Nes, and M. Scheffer, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **370**, 1185 (2012).
- [13] V. Dakos, S. R. Carpenter, E. H. van Nes, and M. Scheffer, *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **370**, 20130263 (2015).
- [14] Y. Sharma, P. S. Dutta, and A. K. Gupta, *Phys. Rev. E* **93**, 032404 (2016).
- [15] C. Perryman and S. Wiczorek, in *Proc. R. Soc. A*, Vol. 470 (The Royal Society, 2014) p. 20140226.
- [16] V. Dakos, M. Scheffer, E. H. van Nes, V. Brovkin, V. Petoukhov, and H. Held, *Proceedings of the National Academy of Sciences* **105**, 14308 (2008).
- [17] P. D. Ditlevsen and S. J. Johnsen, *Geophysical Research Letters* **37** (2010).
- [18] C. Boettiger and A. Hastings, *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20122085 (2012).
- [19] C. Boettiger and A. Hastings, *Proceedings of the Royal Society of London B: Biological Sciences* **280**, 20131372 (2013).
- [20] J. M. Drake, *Proceedings of the Royal Society of London B: Biological Sciences* **280**, 20130686 (2013).
- [21] S. Wiczorek, P. Ashwin, C. M. Luke, and P. M. Cox, in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Vol. 467 (The Royal Society, 2011) pp. 1243–1269.
- [22] H. Risken and T. Frank, *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer Science & Business Media, 2012).
- [23] D. Zhang, G. Wei, D. Kouri, and D. Hoffman, *The Journal of chemical physics* **106**, 5216 (1997).
- [24] G. Williams, *Linear algebra with applications* (Jones & Bartlett Publishers, 2012).
- [25] C. G. Perryman, *How Fast is Too Fast? Rate-induced Bifurcations in Multiple Time-scale Systems*, Ph.D. thesis, University of Exeter (2015).