A Scaling Theory of Bifurcations in the Symmetric Weak-Noise Escape Problem

Robert S. Maier*
rsm@math.arizona.edu
Dept. of Mathematics
University of Arizona
Tucson, AZ 85721, USA

Daniel L. Stein†
dls@physics.arizona.edu
Dept. of Physics
University of Arizona
Tucson, AZ 85721, USA

Abstract

We consider the overdamped limit of two-dimensional double well systems perturbed by weak noise. In the weak noise limit the most probable fluctuational path leading from either point attractor to the separatrix (the most probable escape path, or MPEP) must terminate on the saddle between the two wells. However, as the parameters of a symmetric double well system are varied, a unique MPEP may bifurcate into two equally likely MPEP's. At the bifurcation point in parameter space, the activation kinetics of the system become non-Arrhenius. In this paper we quantify the non-Arrhenius behavior of a system at the bifurcation point, by using the Maslov-WKB method to construct an approximation to the quasistationary probability distribution of the system that is valid in a boundary layer near the separatrix. The approximation is a formal asymptotic solution of the Smoluchowski equation. Our analysis relies on the development of a new scaling theory, which yields 'critical exponents' describing weak-noise behavior near the saddle, at the bifurcation point.

Keywords: boundary layer, caustics, double well, Fokker-Planck equation, Lagrangian manifold, large deviation theory, large fluctuations, Maslov-WKB method, non-Arrhenius behavior, non-generic caustics, Pearcey function, singular perturbation theory, Smoluchowski equation, stochastic escape problem, stochastic exit problem, stochastically perturbed dynamical systems, weak noise, Wentzell-Freidlin theory, WKB approximation.

*Partially supported by the National Science Foundation under grants NCR-90-16211 and DMS-95-00792.
†Partially supported by the U.S. Department of Energy under contract DE-FG03-93ER25155.

1 Introduction

In this paper we build on previous work [28–30] by analysing an unusual bifurcation phenomenon in the theory of noise-activated transitions. We study its appearance in the overdamped limit of two-dimensional double well systems, with nongradient dynamics. In this context, the new phenomenon is a bifurcation of the most probable transition path (in the limit of weak noise) between the two wells, as a system parameter is varied.

In many ways, the behavior of a system whose most probable transition path is just beginning to bifurcate resembles that of a system undergoing a *phase transition*. In particular, double well systems which are 'at criticality' in the bifurcation sense will exhibit non-Arrhenius behavior. This means that the growth of the mean time between inter-well fluctuations, *i.e.*, the growth of the mean time needed for the system to hop from one well to the other, will *not be pure exponential* in the weak-noise limit. In double well systems at criticality, relaxation due to activation will proceed (in the limit of weak noise) at an anomalous, in fact anomalously large, rate.

To treat the previously unnoticed phenomenon of bifurcation, we need to develop a new approach for treating transitions induced by weak noise, when a 'soft mode' appears in the dynamics of transverse fluctuations around the most probable transition path. Since this is analogous to a phase transition, we introduce a *scaling theory*. In the context of double well systems, our scaling theory is a theory of behavior near the saddle point between the two wells, since the saddle is where the most probable inter-well transition path begins to bifurcate. We shall demonstrate that the theory explains the weak-noise behavior, at criticality, of a large universality class of double well systems.

The scaling theory will reveal a striking feature of the bifurcation phenomenon, which is that in any 'critical' double well system there appears (in the weak-noise limit) a *nongeneric singularity* in the stationary probability distribution, located at the saddle point. As Berry [4] discusses, a singularity is nongeneric if it arises, in an appropriate WKB sense, from a catastrophe of unusual type; *i.e.*, one of *infinite codimension*. The stationary distribution near the saddle point is described, in the limit of weak noise, by an unusual (non-canonical) diffraction function. The familiar special functions of WKB theory (Airy functions, Pearcey functions, etc.) do not suffice. The singularity at the saddle, and the diffraction function with which it is 'clothed,' can be viewed as the mathematical source of the non-Arrhenius weak-noise asymptotics.

We begin with three largely qualitative sections. In Section 2 we review the physical relevance of overdamped models with non-gradient dynamics, and in Section 3 explain how the weak-noise behavior of any double well model of this type is determined by its flow field of *instanton trajectories* (most probable fluctuational paths). In Section 4 we sketch the gross features of the bifurcation phenomenon, including features such as further bifurcations and universality. In Section 5, our treatment becomes more quantitative. We first review the matched asymptotic approximations technique we have employed elsewhere [28–30], and begin extending it to handle models with singularities. In Section 5.3 we explain why the bifurcation transition deserves to be called a phase transition. In particular, we explain how behavior near criticality is described by *critical exponents*, which characterize the rate of divergence of measurable quantities (*e.g.*, the pre-exponential factor in the weak-noise asymptotics of the mean inter-well fluctuation time). In Section 5.4 we explain how to determine whether any given double well model is 'critical.' The *transverse Jacobi* *operator* is the differential operator appearing in the second variation of the Onsager-Machlup action functional, when one varies about the most probable inter-well transition path. The onset of bifurcation occurs when this operator acquires a zero eigenvalue.

In Section 6 we explain the use we shall make of Maslov's geometric theory of wave asymptotics [4, 25, 32]. In Section 7 we introduce the concept of a scaling theory, by developing a scaling theory of weak-noise behavior near generic (cusp) singularities. We show how the scaling theory justifies the Ginzburg-Landau approximation used in this context by Dykman *et al.* [12]. In Section 8 we develop an analogous scaling theory for the nongeneric singularity associated with the onset of bifurcation. We compare our theory with numerical data, and examine its predictions for non-Arrhenius behavior and the stationary distribution near the saddle point.

In Section 9 we discuss our results. The reader may wish to glance ahead at Fig. 14, which is an Arrhenius plot of the inter-well hopping rate of any double well system at criticality. The non-Arrhenius behavior shown there, in particular the 'logarithmic bend,' is the key result of this paper.

2 Preliminaries

Statistical physics and chemical physics include many examples of stochastically perturbed dynamical systems. It is often the case that the state of such a system is modelled as a particle moving in an n-dimensional force field $\boldsymbol{F}(\boldsymbol{x})$, and subject to additional random perturbations ('noise'). Since our interest is in the modelling of nonequilibrium systems, we shall not assume (as is usually done) that this force field is conservative.

If the motion of the particle is isotropically damped, with damping constant γ, in the absence of noise the particle position $\boldsymbol{x}$ would obey the deterministic equation

$$m\ddot{\boldsymbol{x}} + \gamma m\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}). \tag{2.1}$$

Adding a random force $\boldsymbol{F}_{\text{random}}(t)$ yields the Langevin equation

$$m\ddot{\boldsymbol{x}} + \gamma m\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}) + \boldsymbol{F}_{\text{random}}(t). \tag{2.2}$$

In physical problems $\boldsymbol{F}_{\text{random}}(t)$ is often modelled as Gaussian white noise with amplitude $\sqrt{2\gamma m k_B T}$, where T is the ambient temperature. In this case the associated partial differential equation, which describes the time evolution of the probability density of $\boldsymbol{x}$ and its velocity, is known as the (forward) Fokker-Planck equation.

A case particularly important in applications is the *overdamped*, or *inertialess* case, when $\gamma \gg t_0^{-1}$, for t_0 the physical time scale. In this case the $m\ddot{\boldsymbol{x}}$ term in (2.2) can be dropped, and the Langevin equation becomes first order in time. If time is rescaled by a factor γm (*i.e.*, $t \leftarrow \gamma m t$), it may be written in the normalized form

$$\dot{\boldsymbol{x}} = \boldsymbol{u}(\boldsymbol{x}) + \epsilon^{1/2}\dot{\boldsymbol{w}}(t). \tag{2.3}$$

Here $\dot{\boldsymbol{w}}(t)$ is a standard n-dimensional Gaussian white noise (the derivative of $\boldsymbol{w}(t)$, a standard n-dimensional Wiener process), the 'drift field' $\boldsymbol{u}$ equals $\boldsymbol{F}$, and ϵ equals $2k_BT$. The corresponding scalar advection-diffusion equation for the probability density $\rho = \rho(\boldsymbol{x}, t)$ of $\boldsymbol{x}$,

$$\dot{\rho} = (\epsilon/2)\nabla^2\rho - \boldsymbol{\nabla} \cdot (\rho\boldsymbol{u}), \tag{2.4}$$

is known as the (forward) Smoluchowski equation. It may be written as $\dot{\rho} = -\mathcal{L}^*\rho$, where

$$\mathcal{L}^* \equiv -(\epsilon/2)\nabla^2 + \boldsymbol{u} \cdot \boldsymbol{\nabla} + \boldsymbol{\nabla} \cdot \boldsymbol{u}. \tag{2.5}$$

It is often necessary to generalize the equation to include the effects of anisotropic damping [21], or state-dependent noise [23]. However, in this paper we consider only overdamped systems whose Langevin equation is of the form (2.3). Since we do not require the deterministic forces to be conservative, *we do not require $\boldsymbol{u}$ to be a gradient field*. This means that even in stationarity, the system may not display detailed balance. Equivalently, the stationary probability distribution for the system may not be (in the traditional sense) in thermal equilibrium.

Attractors of the drift field $\boldsymbol{u}$, in particular point attractors, correspond to 'metastable states': they are stable states of the underlying deterministic dynamics, but the thermal noise may induce transitions between them. Of great physical interest is the time needed for this to occur. For example, how long does it take for the noise in (2.3) to overcome the drift toward a specified stable point S, and drive the system state $\boldsymbol{x}$ beyond the domain of attraction of S, toward another attractor? The study of such noise-activated transitions is known as the *stochastic exit problem*, or the *escape problem*. For general stochastic models only numerical results can be obtained (see, *e.g.*, Ref. [7]). The Smoluchowski equation is particularly difficult to handle in the $\epsilon \to 0$ limit. This is the weak-noise, or low-temperature limit, in which the mean first passage time (MFPT) $\langle\tau\rangle$ from S to the boundary of its domain of attraction grows exponentially. In this limit a single escape path (the most probable escape path, or MPEP) usually dominates. Our approach to the weak-noise limit, which does not rely on a numerical simulation of the Smoluchowski equation, will exploit this asymptotic determinism quite heavily.

3 Symmetric Double Well Models

As in two of our earlier papers on the stochastic exit problem [28, 29], we shall focus on two-dimensional 'double well' systems, with smooth drift field $\boldsymbol{u} = (u_x, u_y)$ of the symmetric form shown in Fig. 1. If $\boldsymbol{x} = (x, y)$ is the two-dimensional state variable, $u_x(x, y)$ is taken to be odd in x and even in y, while for $u_y(x, y)$ the reverse is true. There is assumed to be a linearly stable point attractor $S = (x_s, 0)$ whose domain of attraction is the entire open right-half plane. By symmetry, its reflection $S' = (-x_s, 0)$ attracts the open left-half plane. There is also assumed to be a single *saddle*, or hyperbolic point, on the y-axis separatrix between the two domains of attraction. It must be at the origin, by symmetry. Nongradient drift fields with this topology arise in statistical and chemical physics, and also in theoretical biology, *e.g.*, in stochastic competition models of population dynamics [31].

One expects that as $\epsilon \to 0$, exit from either of the two domains of attraction will occur preferentially over the saddle. The drift field $\boldsymbol{u}$ is assumed to have a nondegenerate linearization at

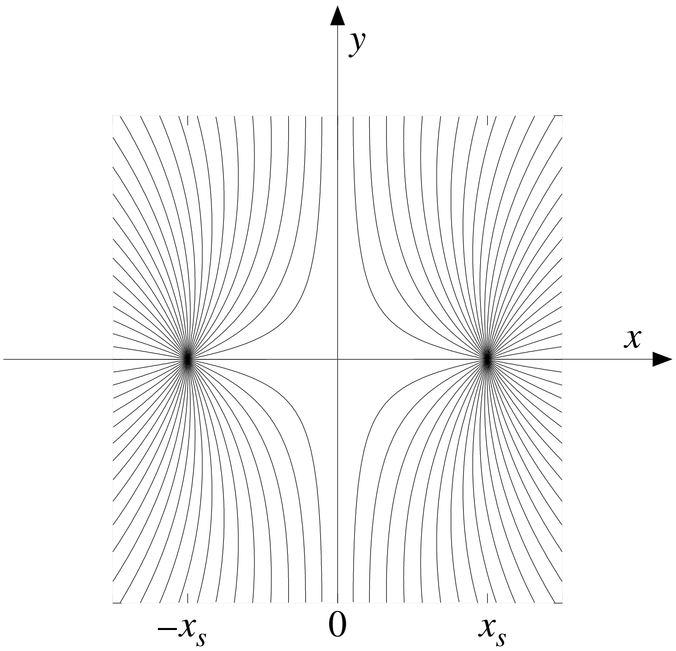

Figure 1: The streamlines of a typical symmetric double well drift field, indicating the path taken by the particle in the absence of noise. There are point attractors at $S = (x_s, 0)$ and $S' = (-x_s, 0)$, and a saddle point at $(0, 0)$.

the saddle. So $\lambda_x = \partial u_x/\partial x(0,0) > 0$, and $\lambda_y = \partial u_y/\partial y(0,0) < 0$. We shall see that the character of the abovementioned bifurcation phenomenon depends strongly on the quotient $\mu \equiv |\lambda_y|/\lambda_x$.

A typical (and not necessarily gradient) symmetric double well drift field, which we have used elsewhere for purposes of illustration and shall examine further below, is

$$\begin{aligned} u_x(x,y) &= x - x^3 - \alpha x y^2, \\ u_y(x,y) &= -\mu(1+x^2)y, \end{aligned} \tag{3.1}$$

in which μ appears as a parameter. We shall call this drift field the 'standard' double well model. For any choice of $\mu > 0$, its structure is that of Fig. 1, with $S = (x_s, 0) = (1, 0)$. It is not a gradient field unless the parameter α equals μ. If $\alpha > 0$ it has a very significant additional property, which we shall require of all our double well models. This is the property that $\partial^2 u_x/\partial y^2(x, 0)$, which by symmetry is an odd function of x, is strictly negative for all x between 0 and x_s. If this is the case, the drift from the saddle toward S 'softens' as one moves away from the x-axis. The off-axis softening, for the standard model, increases as α is increased.

We remind the reader of our approach to the weak-noise limit of stochastically perturbed dynamical systems. (We review the mathematical aspects in Section 5.1 and 5.2.) Suppose that such a system has a unique stationary probability density ρ_0, which satisfies the time-independent Smoluchowsk equation $\mathcal{L}^*\rho_0 = 0$. Typically, as the noise strength $\epsilon \to 0$, ρ_0 takes on an asymptotic WKB form. In fact

$$\rho_0(\boldsymbol{x}) \sim K(\boldsymbol{x}) \exp[-W(\boldsymbol{x})/\epsilon], \quad \epsilon \to 0, \tag{3.2}$$

for certain functions W and K whose smoothness properties we shall leave unspecified; K, in particular, may have singularities. In any double well model, by convention $W = 0$ and $K = 1$ at $\boldsymbol{x} = S$ and S'. Moreover, $W > 0$ at all points $\boldsymbol{x}$ other than S and S'. W is called the *nonequilibrium potential* of the model [16]. If the drift $\boldsymbol{u}$ equalled the negative gradient of a potential $\boldsymbol{\Phi}$, then W would equal $2\boldsymbol{\Phi}$, K would reduce to a constant, and the WKB form (3.2) would reduce to a Maxwell-Boltzmann distribution. For systems with nongradient dynamics, the computation of W and K is more complicated.

In general W has an alternative interpretation as a *classical action function*. As we review in Section 5.1, this is because the WKB approximation (3.2) is determined by a flow field of 'classical' trajectories, or WKB characteristics, emanating from the attractors of the deterministic dynamics (*e.g.*, S and S'). These classical trajectories (sometimes called *instanton* trajectories [6, 28], or *optimal* trajectories [12]) have a physical interpretation as *most probable fluctuational paths*. In the double well case, the exponentially rare fluctuations from S (resp. S') to any point $\boldsymbol{x}$ in its domain of attraction become increasingly concentrated around the classical trajectory extending from S (resp. S') to $\boldsymbol{x}$. Equivalently, the most probable 'prehistory' of any fluctuation passing through $\boldsymbol{x}$ extends back toward S or S' along this trajectory [11]. The trajectories are determined by a classical Lagrangian (the Onsager-Machlup Lagrangian), and $W(\boldsymbol{x})$ is obtained by integrating this Lagrangian along the classical trajectory terminating at $\boldsymbol{x}$. $W(\boldsymbol{x})$ is interpreted as the rate at which fluctuations to the neighborhood of $\boldsymbol{x}$ are suppressed exponentially, as $\epsilon \to 0$.

In symmetric double well models the stationary density ρ_0 (and hence W) must be even in x. In the $\epsilon \to 0$ limit the phenomenon of noise-activated hopping between the two wells is governed by the closely related *quasistationary* density ρ_1, which is odd rather than even. The quasistationary density is the next lowest lying (*i.e.*, slowest decaying) eigenmode of the Smoluchowski operator $\mathcal{L}^*$. For any choice of initial conditions the probability density $\rho = \rho(\boldsymbol{x}, t)$ necessarily satisfies

$$\rho(\boldsymbol{x}, t) \sim \rho_0(\boldsymbol{x}) + C\rho_1(\boldsymbol{x}) \exp(-\lambda_1 t), \quad t \to \infty \tag{3.3}$$

for some constant C, where λ_1 is the eigenvalue of ρ_1. The exponential decay of the quasistationary eigenmode is interpreted as describing the equilibration of probability between the two wells due to noise-activated hopping, or the absorption of probability on the separatrix [7]. ρ_1 of course satisfies Dirichlet (absorbing) boundary conditions on the separatrix. Its eigenvalue $\lambda_1 = \lambda_1(\epsilon)$ normally falls to zero exponentially as $\epsilon \to 0$. The exponentially small splitting between the ground state eigenvalue $\lambda_0 \equiv 0$ and the eigenvalue λ_1 is analogous to the exponentially small splitting (as $\hbar \to 0$) between the ground state and first excited state of a quantum-mechanical Hamiltonian with double well potential. Both are WKB phenomena. In the $\epsilon \to 0$ limit, λ_1 is interpreted as the rate at which noise-activated hopping takes place. Equivalently, it is a *reciprocal MFPT*.

The techniques reviewed in Section 5.1 and 5.2 permit a computation of the $\epsilon \to 0$ asymptotics of the eigenvalue λ_1, and hence of the MFPT $\langle\tau\rangle$, in most symmetric double well models. Our basic approach is similar to that of Kramers [22]. In the limit of weak noise we approximate $\rho_1(x, y)$ by $\rho_0(x, y)\,\mathrm{sgn}(x)$ except in a 'boundary layer' of width $O(\epsilon^{1/2})$ near the $x = 0$ separatrix, and compute λ_1 as the rate at which probability is absorbed on the separatrix. Performing this computation requires the construction of a boundary layer approximation to ρ_1, valid near the saddle, and matching to the 'outer' approximations on either side [8]. Normally, we find $\langle\tau\rangle \sim A \exp[+W(0, 0)/\epsilon]$, where $A \propto K(0, 0)^{-1}$. So the asymptotic MFPT growth rate in the limit of weak noise is simply $\Delta W = W(0, 0) - W(S) = W(0, 0)$, the height of the 'action barrier,' or activation barrier, between the two wells. And the MFPT generally displays a pure exponential (Arrhenius) growth, with an explicitly computable (ϵ-independent) prefactor. We shall see, however, that the bifurcation phenomenon may induce more complicated (non-Arrhenius) weak-noise asymptotics for the MFPT.

4 The Bifurcation Phenomenon: Qualitative Features

We pointed out in Ref. [28] that a bifurcation phenomenon may occur in double well models as their parameters are varied. Figure 2 displays the flow of instanton trajectories (*i.e.*, most probable weak-noise fluctuational paths) emanating from the stable point $S = (1, 0)$ in the standard model (3.1) with $\mu = 1$, at several values of the parameter α. When $0 < \alpha < 4$ the general picture resembles Fig. 2(a): the line segment from $(1, 0)$ to the saddle $(0, 0)$ is the only instanton trajectory from S to the saddle. This line segment is interpreted as the most probable escape path (MPEP). In the weak-noise limit, the (exponentially rare) fluctuations from the right-half plane to the left-half plane proceed preferentially along it. To leading order, activation kinetics reduce to instanton dynamics.

As α is increased, there is a qualitative change, akin to a phase transition, in the behavior of the instanton trajectories. This takes place at the critical value $\alpha = 4$, as shown in Fig. 2(b). When $\alpha > 4$ as in Fig. 2(c), they *focus* at a point $(x_f, 0)$ on the x-axis, with $x_f > 0$. x_f converges to

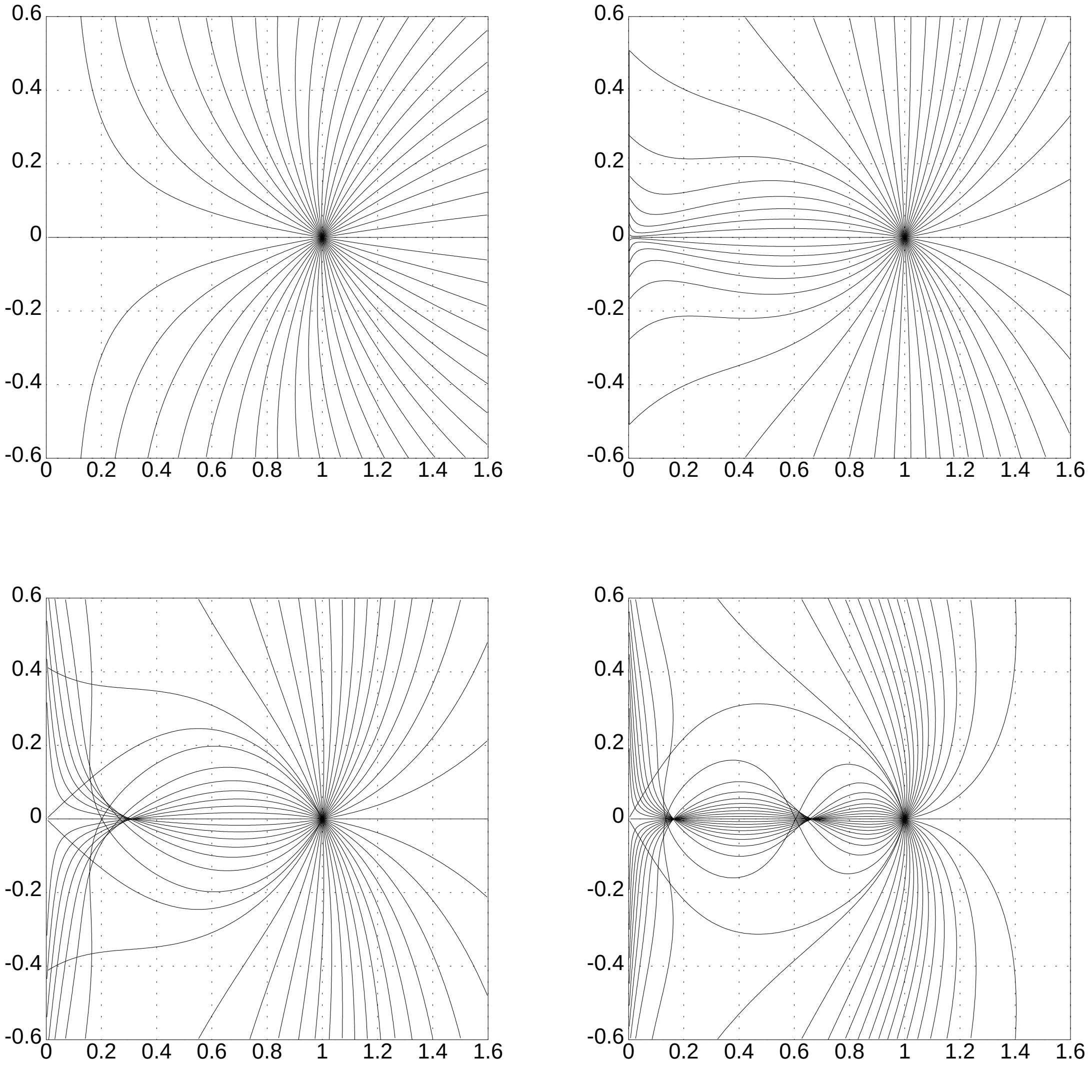

Figure 2: The flow field of outgoing instanton trajectories (*i.e.*, most probable fluctuational paths, in the weak-noise limit) emanating from the stable point $S = (1, 0)$ of the standard double well model (3.1). Here $\mu = 1$, and parts (a), (b), (c), (d) of the figure illustrate the cases $\alpha = 1, 4, 5, 10$. The $\alpha = 4$ model is 'critical' in the bifurcation sense. Increasing α above 4 causes the instanton trajectories to focus, and the MPEP to bifurcate.

zero as $\alpha \to 4^+$, so one may speak of the focal point 'being born' at criticality, and 'emerging from the saddle' as α is increased above its critical value. In geometrical optics the focal point would be called a *cusp*. From it there extends a *fold*, or *caustic* (an envelope of crossing trajectories, with $\Delta y \propto (\Delta x)^{3/2}$).

Each point in the sharp-tipped region within the fold is reachable from S via *three* instanton trajectories. Of these three trajectories, only the one(s) with minimum action are 'physical,' and can be interpreted as most probable fluctuational paths. For example, on-axis points $(x, 0)$ with $0 \le x < x_f$ are reachable via an on-axis (straight) trajectory, and via two additional symmetrically placed off-axis (curved) trajectories. Computation shows that the off-axis trajectories have lesser action, and are dominant. The true ('least action') MPEP's in Fig. 2(c) are accordingly the symmetrically placed pair of off-axis trajectories, one above and one below the x-axis, that terminate on the saddle. Note that beyond the cusp (*i.e.*, at $x < x_f$), the physical action W is no longer differentiable through the x-axis. This nondifferentiability arises from different dominant off-axis trajectories being selected as $y \to 0^+$ and $y \to 0^-$.

The transition at $\alpha = 4$ can be interpreted as a *bifurcation of the MPEP*, corresponding to a sort of symmetry breaking. At larger values of α, the drift field $\boldsymbol{u}$ and the Langevin equation (2.3) remain symmetric about the x-axis, but each of the two MPEP's is not. The line segment from S to the saddle, formerly the (unique) MPEP, in no way contributes to the leading weak-noise asymptotics for escape. (It remains an extremum of the Onsager-Machlup action functional, but is no longer the minimum.)

The occurrence of a bifurcation in the standard model at sufficiently high α (when $\mu = 1$, at $\alpha = 4$) is due to the fact that by increasing α, one softens the resistance to motion toward the separatrix in the vicinity of the x-axis (though not on the x-axis itself). This enhances the probability of escape trajectories that deviate from the axis. Of course it is only in the limit, as $\epsilon \to 0$, that well-defined MPEP's appear. And the existence of a sharp, well-defined transition when α equals some critical value α_c is not at all obvious!

When α is increased beyond α_c, further bifurcations of the on-axis instanton trajectory will occur. In Section 5.4 we explain how the critical values of α are determined by a *Jacobi equation*, with a classical mechanical interpretation. It turns out that in the standard model with $\mu = 1$ the j'th bifurcation occurs at $\alpha = \alpha_c^{(j)} = (j + 1)^2$. Figure 2(d) shows the situation at $\alpha = 10$, when a second focus $(x_f^{(2)}, 0)$ has emerged, with its own caustic. Beyond the first focus $(x_f, 0)$ each point on the x-axis is reached from S by *three* instanton trajectories; beyond the second focus, each such point is reached by *five*. The MPEP's in Fig. 2(d), however, remain the symmetrically placed pair of off-axis trajectories that terminate on the saddle. Computation shows that the oscillatory trajectories from S to the saddle arising from the second, third,... bifurcations have higher actions, and are accordingly not physical.

That caustics can occur in the flow pattern of the most probable fluctuational paths has been known for some time [9, 19], but our Ref. [28] was the first to consider the effects on exit phenomena. We shall see that what occurs at the first critical value of α has much in common with a critical point characterizing a *phase transition* in a condensed matter system. This is suggested by Fig. 3, which plots the activation barrier $\Delta W = W(0, 0)$, as determined by the true MPEP or MPEP's, as a function of α for the standard model with $\mu = 1$. (Recall that $\Delta W = W(0, 0)$ is the exponential

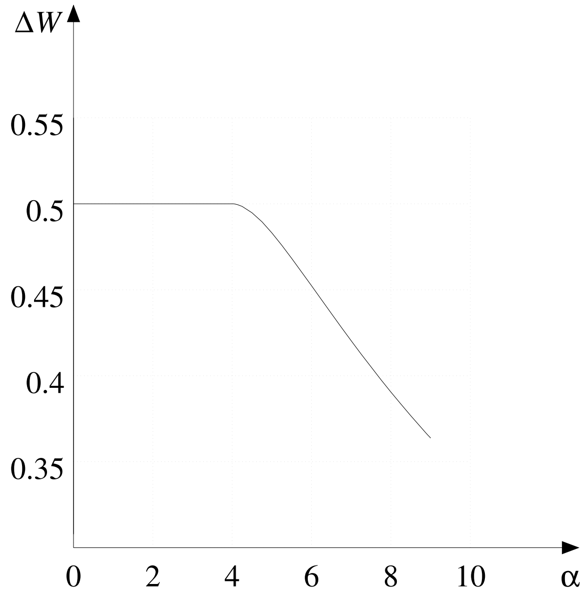

Figure 3: The activation barrier $\Delta W = W(0,0) - W(S) = W(0,0)$ between the two wells of the standard double well model (3.1), as a function of the off-axis softening parameter α. Here $\mu = 1$. The lowering of the activation barrier beyond $\alpha = \alpha_c = 4$ is due to the bifurcation of the MPEP, along which the action difference ΔW is computed.

growth rate of the MFPT as the noise strength tends to zero.) $W(0,0)$ decreases above $\alpha = \alpha_c = 4$ as the bifurcating MPEP's move away from the x-axis. The WKB prefactor $K(0,0)$ turns out to be singular at the bifurcation transition; in Section 5.3 we note that it diverges as $\alpha \to \alpha_c^-$. As a consequence, in the standard model at least, *the weak-noise MFPT asymptotics at criticality cannot be of a pure Arrhenius form*.

There is in fact a set of critical exponents describing the behavior of the $\epsilon \to 0$ asymptotics of the standard model as α tends to its (μ-dependent) first critical value $\alpha_c = \alpha_c^{(1)}$, and as $x \to 0$, $y \to 0$. It is a reasonable conjecture that behavior near criticality is *universal* in the sense that it does not depend on the details of the stochastic model exhibiting the bifurcation phenomenon. To analyse the critical behavior and demonstrate universality, in Sections 7 and 8 we begin the construction of a *scaling theory* of the bifurcation phenomenon. Our treatment extends from the standard model (3.1) to any symmetric double well model with a similar 'off-axis softening parameter' α, and a first critical value α_c. We first identify the singular behavior, for any double well model at criticality, of the action W and the WKB prefactor K at the saddle point. We then show that at criticality, the stationary density ρ_0 and the quasistationary density ρ_1 may be approximated on an appropriate (ϵ-dependent) length scale near the saddle point by certain 'diffraction functions,' which have explicit integral representations. The technique for constructing these representations is due to Maslov [32], and ultimately to Keller [20]. It was Maslov who first worked out, in the context of wave fields, the diffraction functions that 'clothe' generic singularities other than cusps and folds.

A very important discovery, from a mathematical point of view, will be that when α equals the critical value α_c where the MPEP begins to bifurcate, the saddle point $(0,0)$ acquires a certain nonzero *singularity index*. What this means is best understood by comparing the singularity at the saddle (when $\alpha = \alpha_c$) with the cusp and fold singularities present when $\alpha > \alpha_c$. The terminology of geometrical optics [4] is appropriate. The cusp at $(x_f, 0)$ is a structurally stable singularity (or *catastrophe*, in the language of Thom), with codimension 2. The fold extending from it, though not 'physical' in the above least-action sense, is a catastrophe of codimension 1. For points $\boldsymbol{x}$ in the vicinity of the cusp, the WKB approximation (3.2) for the value of the stationary density $\rho_0(\boldsymbol{x})$ breaks down. The proper treatment of points near the cusp and the fold is similar to the short-wavelength treatment of wave fields near caustics [4, 12]. The cusp is said to have singularity index $1/4$, and points on the fold would (if it were physical) have singularity index $1/6$. This means that at these singular points the prefactor in the WKB approximation to ρ_0, which formally diverges, if properly constructed would acquire a factor $\epsilon^{-1/4}$ (resp. $\epsilon^{-1/6}$). There is a non-WKB (but uniformly valid) approximation to $\rho_0(\boldsymbol{x})$ in the vicinity of each such singular point, in terms of canonical diffraction functions. We shall re-derive these facts in Section 7, in terms of scaling functions.

We shall show in Section 8 that the singularity index of the point singularity appearing at the saddle, in critical models, depends in a universal way on μ, *i.e.*, on the ratio of the eigenvalues of the linearization of the drift $\boldsymbol{u}$ at the saddle. It turns out to equal $(\mu+1)/6$. Moreover, the approximations to ρ_0 and ρ_1 near the saddle are given by *non-canonical* diffraction functions. By using the non-canonical approximation to ρ_1 to compute the rate at which probability is absorbed on the separatrix we shall quantify the *universal non-Arrhenius behavior* of the weak-noise MFPT

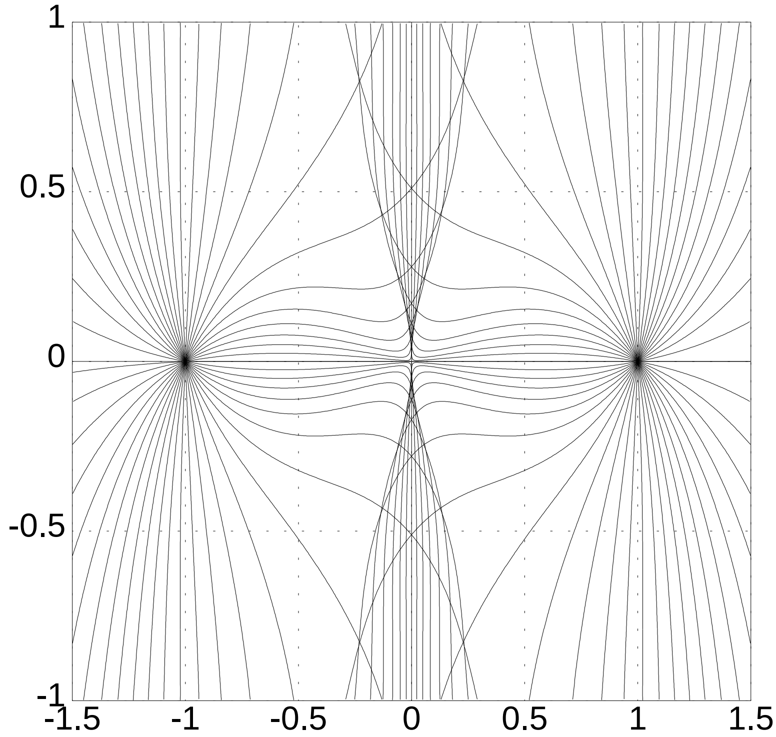

Figure 4: The flow field of the instanton trajectories emanating from both stable fixed points S and S', in a critical version of the standard double well model (3.1). This figure reveals that at criticality, a two-sided caustic extends sideways from the saddle point. Although a universal phenomenon, this caustic is nongeneric in the sense of singularity theory. Here $\mu = 1$, and the parameter α is set equal to the corresponding first critical value $\alpha_c = 4$, as in Fig. 2(b).

asymptotics. We shall show that in symmetric double well models at criticality,

$$\langle \tau \rangle \sim \text{const} \times \epsilon^s \exp[+\Delta W/\epsilon], \quad \epsilon \to 0, \tag{4.1}$$

where $s = s(\mu) = (\mu + 1)/6$ is the index of the singularity at the saddle. We shall also derive scaling corrections, in the weak-noise limit, to the normal distribution of exit location points near the saddle. The preceding results will hold for all μ satisfying $3/4 < \mu < 3$; the weak-noise asymptotics of models with $\mu \leq 3/4$ and $\mu \geq 3$ are still under investigation.

The point singularity appearing in critical models at the saddle point, which may be termed a *nascent cusp*, is *nongeneric*. It is not a member of the well known family of singularities that includes folds, cusps, swallowtails, etc. This becomes clear if one plots the flow field of the instanton trajectories emanating from both S and S' in the standard model at criticality (μ varying, and α set equal to its μ-dependent first critical value). At least when $3/4 < \mu < 3/2$, one finds that at criticality *a two-sided caustic extends transversally from the saddle point itself*. (Cf. Dykman *et al.* [12].) Figure 4, which is an extended version of Fig. 2(b), shows the flow field when $\mu = 1$

and $\alpha = \alpha_c = 4$. The caustic is clearly visible. It is not 'physical,' since it is formed by high-action instanton trajectories that have crossed the separatrix. But the 'nascent' cusp is clearly a cusp in its own right, of an unusual sort. Numerically one finds that the two-sided caustic extending from it is located at

$$|x| \lesssim \text{const} \times |y|^{(3/2-\mu)^{-1}}, \quad y \to 0. \tag{4.2}$$

A conventional (generic) caustic would have an exponent of $3/2$. The continuously varying exponent $(3/2 - \mu)^{-1}$, which turns out to be universal and which we shall derive in Section 8 from our scaling theory, signals that the two-sided caustic is nongeneric. The nascent cusp from which it extends is itself nongeneric in the sense of singularity theory.

As Berry [4] has emphasized, nongeneric singularities arise from *catastrophes of infinite codimension*. It is remarkable that a singularity of such complexity is a universal feature of singly parametrized symmetric double well models with non-gradient dynamics.

5 Quantitative Semiclassical Asymptotics

We now begin a quantitative treatment of the weak-noise asymptotics for escape. We first recast our earlier results in a form that facilitates the analysis of singularities. In Section 5.1 we discuss geometric aspects of the WKB approximation, and in Section 5.2 we discuss our matched asymptotic approximations technique for computing MFPT asymptotics. In Section 5.3 we use the standard model (3.1) to illustrate the nature of the nascent cusp appearing at the saddle at criticality, and the ways in which bifurcation can be viewed as a phase transition. In Section 5.4 we explore the bifurcation phenomenon from a classical mechanical point of view, and relate it to the appearance of a *transverse soft mode*. We explain how its appearance is governed by a Jacobi equation, and how this equation determines whether or not a given double well model is at criticality.

5.1 The WKB Approximation and Classical Mechanics

The time-independent forward Smoluchowski equation $\mathcal{L}^*\rho_0 = 0$ may be written as

$$H(\boldsymbol{x}, -\epsilon\boldsymbol{\nabla})\rho_0 = 0 \tag{5.1}$$

where

$$H(\boldsymbol{x}, \boldsymbol{p}) = \boldsymbol{p}^2/2 + \boldsymbol{u}(\boldsymbol{x}) \cdot \boldsymbol{p} \tag{5.2}$$

is the so-called Wentzell-Freidlin Hamiltonian [14], whose dual is the Onsager-Machlup Lagrangian [34]

$$L(\boldsymbol{x}, \dot{\boldsymbol{x}}) = |\dot{\boldsymbol{x}} - \boldsymbol{u}(\boldsymbol{x})|^2/2. \tag{5.3}$$

In equation (5.1) we have adopted an operator ordering convention according to which the action of $\boldsymbol{\nabla}$ precedes that of $\boldsymbol{x}$.

In the weak-noise ($\epsilon \to 0$) limit the stationary density ρ_0 and the quasistationary density ρ_1 are given in the interior of each well by a WKB, or semiclassical form. A full WKB expansion for ρ_0 would be of the form

$$\rho_0(\boldsymbol{x}) \sim [K^{(0)}(\boldsymbol{x}) + \epsilon K^{(1)}(\boldsymbol{x}) + \cdots] \exp[-W(\boldsymbol{x})/\epsilon], \qquad \epsilon \to 0, \tag{5.4}$$

as in geometrical optics. By substituting this formal series into (5.1) and examining the coefficients of each power of ϵ, one obtains equations for W and the $K^{(m)}$. That is what we shall do, though we shall work only to leading order: our WKB Ansatz will be $\rho_0(\boldsymbol{x}) \sim K(\boldsymbol{x}) \exp[-W(\boldsymbol{x})/\epsilon]$. Notice that since the eigenvalue $\lambda_1 = \lambda_1(\epsilon)$ of ρ_1 is exponentially small as $\epsilon \to 0$, the asymptotic expansions (in powers of ϵ) for ρ_1 and ρ_0 will be the same. To see the difference between them, which is significant only near the separatrix between the two wells, one would have to go 'beyond all orders' in the WKB expansion.

The eikonal equation for W is the time-independent Hamilton-Jacobi equation

$$H(\boldsymbol{x}, \boldsymbol{\nabla} W) = 0 \tag{5.5}$$

so that W is a *classical action at zero energy*. For any point $\boldsymbol{x}$ in either well, it may be computed by integrating the Lagrangian along the zero-energy classical trajectory extending from S (resp. S') to $\boldsymbol{x}$. Each such trajectory, which satisfies the Euler-Lagrange equations, is interpreted as a most probable fluctuational path in the $\epsilon \to 0$ limit. These trajectories are the 'instanton trajectories' of the last section; the term is justified by analogy with the semiclassical limit in quantum mechanics and quantum field theory [3]. In the language of Gutzwiller [18], the points at which the instanton trajectories focus would be called *zero-energy conjugate points.*

It is convenient to work in the Hamiltonian picture, according to which the classical trajectories of interest lie on a zero-energy surface in a nonphysical phase space, coordinatized by position $\boldsymbol{x}$ and momentum $\boldsymbol{p}$. The flow in this phase space ($2n$-dimensional, if configuration space is n-dimensional) is determined by Hamilton's equations and the Wentzell-Freidlin Hamiltonian. From this point of view the instanton trajectories of Figs. 2 and 4 are mere images of phase-space trajectories, projected 'down' to configuration space by the map $(\boldsymbol{x}, \boldsymbol{p}) \mapsto \boldsymbol{x}$. The phase-space trajectories emanate from $(S, \mathbf{0})$ (resp. $(S', \mathbf{0})$). In WKB theory the projected trajectories are traditionally called *characteristics*, and the phase space trajectories *bicharacteristics*. Characteristics may intersect, as in Figs. 2(c) and 2(d), but bicharacteristics may not.

It is easy to verify, using Hamilton's equations, that $(S, \mathbf{0})$ and $(S', \mathbf{0})$ are hyperbolic fixed points of the Hamiltonian flow. And the unstable manifold of $(S, \mathbf{0})$, for example, comprises all points $(\boldsymbol{x}, \boldsymbol{p})$ that lie on one of the bicharacteristics emanating from $(S, \mathbf{0})$. The unstable manifolds of $(S, \mathbf{0})$ and $(S', \mathbf{0})$ are *Lagrangian* [25]: they are invariant under the Hamiltonian flow. By the term 'Lagrangian manifold' we shall refer to either of these two unstable manifolds, or their union. We denote by $\mathcal{M}$ this union, *i.e.*, the set of all points $(\boldsymbol{x}, \boldsymbol{p})$ that lie on a bicharacteristic emanating from either $(S, \mathbf{0})$ or $(S', \mathbf{0})$. If configuration space is n-dimensional, $\mathcal{M}$ will be an n-dimensional manifold.

Each point $\mathfrak{p} = (\boldsymbol{x}, \boldsymbol{p})$ on $\mathcal{M}$ has a value for the zero-energy action W associated with it, computed by

$$W(\mathfrak{p}) = \int \boldsymbol{p} \cdot d\boldsymbol{x}, \tag{5.6}$$

the line integral being taken along the bicharacteristic terminating at $\mathfrak{p}$. If due to intersecting characteristics, or the crossing of characteristics from one well to the other, there are several manifold points $\mathfrak{p}_i = (\boldsymbol{x}, \boldsymbol{p}^{(i)})$ 'above' some point $\boldsymbol{x}$, then $W(\boldsymbol{x})$ and its gradient $\boldsymbol{p} = \boldsymbol{p}(\boldsymbol{x})$ will in a mathematical sense be multivalued. As a function of $\boldsymbol{x}$, W may in fact have branch points, branch lines (cuts), etc. But the *physical* action $W(\boldsymbol{x})$ appearing in the WKB approximation will be single-valued: it will equal the minimum of the values $W(\mathfrak{p}_i)$ at the manifold points above $\boldsymbol{x}$. This 'least action' computation determines which instanton trajectories are physical.

The WKB prefactor K satisfies an easily derived transport equation. (Cf. Talkner [38].) If one uses the fact that $\dot{\boldsymbol{x}} = \boldsymbol{p} + \boldsymbol{u}(\boldsymbol{x})$ (which is one of Hamilton's equations), the transport equation takes on the comparatively simple form

$$\dot{K} = -[\boldsymbol{\nabla} \cdot \boldsymbol{u} + \nabla^2 W/2]K, \tag{5.7}$$

the time derivative referring to instanton transit time, *i.e.*, to motion along a characteristic or bicharacteristic. Similarly to W, K may be regarded as a function on $\mathcal{M}$ rather than on configuration space. Integration of the equation (5.7) requires knowledge of the second spatial derivatives of W along the characteristic. But $(\partial^2 W/\partial x_i \partial x_j)(\boldsymbol{x})$ equals $(\partial p_i/\partial x_j)(\boldsymbol{x})$, which is a measure of the 'slope' of the manifold above the point $\boldsymbol{x}$. By differentiating Hamilton's equations it is easy to show that the Hessian matrix $\boldsymbol{Z} = (Z_{ij})$ whose elements are the partial slopes $\partial p_i/\partial x_j$ satisfies the matrix Riccati equation

$$\dot{\boldsymbol{Z}} = -\boldsymbol{Z}^2 - \boldsymbol{Z}\boldsymbol{B} - \boldsymbol{B}^t\boldsymbol{Z} - \sum_l p_l \boldsymbol{Y}^{(l)} \tag{5.8}$$

along any characteristic. (Cf. Ludwig [26].) Here $\boldsymbol{B} = (\partial u_i/\partial x_j)$ and $\boldsymbol{Y}^{(l)} = (\partial u_l/\partial x_i \partial x_j)$ are auxiliary matrices. Since $\nabla^2 W = \operatorname{tr} \boldsymbol{Z}$, the computation of K by numerical integration is straightforward.

It is interesting to compare these results with those of Littlejohn [25] on the WKB prefactor for the solutions of the Schrödinger equation in the semiclassical ($\hbar \to 0$) limit. He introduces a Lagrangian manifold, and a similar integration along characteristics. But because he analyses the *time-dependent* Schrödinger equation, he finds that the transport equation for his analogue of K can be integrated explicitly, yielding a Van Vleck determinant. Matters are not so simple in the time-independent case, for the Schrödinger equation as well as for the Smoluchowski equation. Our WKB analysis of the weak-noise limit of the stationary density actually has more in common with the work of Gutzwiller on the semiclassical approximation of fixed-energy quantum-mechanical Green's functions [17, 18, 36] than it does with the semiclassical approximation of time-dependent quantum-mechanical propagators. The prefactor K is analogous to the prefactor of a semiclassical Green's function (at fixed energy). It can in fact be related to the *density of bicharacteristics* on the Lagrangian manifold. This resembles Gutzwiller's interpretation of the prefactor of a semiclassical Green's function in terms of the density of classical trajectories on an energy surface [18].

5.2 Matched Asymptotic Approximations and the MFPT

We now specialize to two-dimensional double well models with the structure of Fig. 1. On account of symmetry and smoothness we may expand the drift $\boldsymbol{u} = (u_x, u_y)$ thus:

$$\begin{aligned} u_x(x,y) &= v_0(x) + v_2(x)y^2 + \cdots \\ u_y(x,y) &= u_1(x)y + u_3(x)y^3 + \cdots . \end{aligned} \tag{5.9}$$

By assumption $v_0(x) > 0$ for all x between 0 and x_s, and $u_1(x) < 0$ for all x between 0 and x_s inclusive. If the symmetry through the axis is unbroken, W and K (both of them computed by integration along instanton trajectories emanating from S) will have similar expansions

$$W(x,y) = w_0(x) + w_2(x)y^2/2! + \cdots \tag{5.10}$$

$$K(x,y) = k_0(x) + k_2(x)y^2/2! + \cdots . \tag{5.11}$$

Here $w_{2m}(x) \equiv \partial^{2m}W/\partial y^{2m}(x,0)$ and $k_{2m}(x) \equiv \partial^{2m}K/\partial y^{2m}(x,0)$. Since W can be viewed as a classical action, the functions w_{2m} can be expressed in terms of the momentum $\boldsymbol{p} = \boldsymbol{p}(\boldsymbol{x})$ of the instanton trajectories passing through near-axis points $\boldsymbol{x}$. For example, $w_0'(x) = p_x(x,0)$ and $w_2(x) = \partial p_y/\partial y(x,0)$. Substituting the WKB Ansatz into the Smoluchowski equation $\mathcal{L}^*\rho_0 = 0$, and examining the coefficients of each power of ϵ and y, will yield equations for the various coefficient functions in (5.10) and (5.11). One finds in particular that $w_0' = p_x = -2v_0$, or

$$w_0(x) = 2\int_x^{x_s} v_0(x')\,dx'. \tag{5.12}$$

Therefore the Hamilton equation $\dot{x} = p_x + v_0(x)$, which follows from the Wentzell-Freidlin Hamiltonian (5.2), implies that $\dot{x}$ must equal $-v_0(x)$ at all points between S and the saddle. The instanton trajectory on the x-axis moves with a speed equal to the local value of the drift speed, but in the direction opposite to the drift.

Examining coefficients also yields the two equations

$$\dot{k}_0 = -[u_1 + w_2/2]k_0 \tag{5.13}$$

$$\dot{w}_2 = -w_2^2 - 2u_1w_2 + 4v_0v_2 \tag{5.14}$$

where we have changed the independent variable from x to t by writing $\dot{k}$ for $-v_0k_0'$, and $\dot{w}_2$ for $-v_0w_2'$. Equations (5.13) and (5.14) could equally well be deduced from (5.7) and (5.8). For later reference we note that

$$\begin{aligned} \dot{w}_4 &= -v_0w_4' \\ &= -4(w_2+u_1)w_4 - 3[(w_2')^2 + 4v_2w_2' + 8u_3w_2] + 48v_0v_4 \end{aligned} \tag{5.15}$$

is the equation satisfied by the fourth derivative $w_4 = \partial^4W/\partial y^4 = \partial^3p_y/\partial y^3$ on the x-axis.

The physical interpretation of the functions k_0 and w_2 is straightforward. The WKB Ansatz implies that

$$\rho_0(x,y) \sim k_0(x)\exp\left\{-\left[w_0(x) + w_2(x)\frac{y^2}{2!}\right]/\epsilon\right\}, \quad \epsilon \to 0. \tag{5.16}$$

So when ϵ is small, w_2 governs the *small transverse fluctuations* about the x-axis. At any time when the system state $\boldsymbol{x}$ has fluctuated leftward from x_s to x, the distribution of the transverse component y will (provided that $w_2(x) > 0$) be approximately Gaussian, with variance $\sim \epsilon/w_2(x)$. Of course such fluctuations are exponentially rare, on account of the $\exp[-w_0(x)/\epsilon]$ factor.

The Riccati equation (5.14) therefore gives the position dependence of the width of the 'WKB tube' of probability density surrounding the MPEP, when this MPEP is in fact the line segment between S and the saddle. Moreover this equation captures the essence of the bifurcation phenomenon, as we shall see in Section 5.4. For the moment we note only that it may readily be integrated from $t = -\infty$ (when the instanton trajectory formally emerges from S) to $t = +\infty$ (when the trajectory, obeying $\dot{x} = -v_0(x)$, reaches the saddle). Since $v_0(x(t)) \to 0$ as $t \to \pm\infty$, we see from (5.14) that w_2 must converge as $t \to +\infty$ (resp. $t \to -\infty$) to one of the two zeroes of the quadratic polynomial

$$-w_2^2 - 2u_1 w_2 = -w_2(w_2 - 2|u_1|), \tag{5.17}$$

where u_1 signifies $u_1(0)$ (resp. $u_1(x_s)$). On physical grounds one expects that usually ('generically') the WKB tube will have a finite variance at both endpoints, *i.e.*, as $t \to -\infty$ and $x \to x_s$, and as $t \to +\infty$ and $x \to 0$. So $w_2(x_s)$ should equal $2|u_1(x_s)|$, and $w_2(0)$ should equal $2|u_1(0)|$.

If these endpoint ('turning point') conditions hold, it is easy to match the tube approximation (5.16) to auxiliary, non-WKB approximations valid near the endpoints: the stable points and the saddle. On physical grounds, ρ_0 and ρ_1 may be approximated on the $O(\epsilon^{1/2})$ lengthscale near S by a Gaussian function of the system state $\boldsymbol{x}$. Let us write ν_x and ν_y for $\partial u_x/\partial x(S)$ and $\partial u_y/\partial y(S)$, the two (negative) eigenvalues of the linearization of the drift $\boldsymbol{u}$ at S. Then

$$\rho_0(x,y) \sim \text{const} \times e^{-|\nu_x|(x-x_s)^2/\epsilon} e^{-|\nu_y|y^2/\epsilon}, \quad \epsilon \to 0 \tag{5.18}$$

near S, the same being true of ρ_1. Since $\nu_y = u_1(x_s)$, this will match to the tube approximation if $w_2(x_s) = 2|u_1(x_s)|$. Similarly, on the $O(\epsilon^{1/2})$ lengthscale near the saddle, ρ_0 may be approximated by the inverted Gaussian

$$\rho_0(x,y) \sim \text{const} \times e^{+\lambda_x x^2/\epsilon} e^{-|\lambda_y|y^2/\epsilon}, \quad \epsilon \to 0. \tag{5.19}$$

Since $\lambda_y = u_1(0)$, the tube approximation will match to (5.19) if $w_2(0) = 2|u_1(0)|$ and $k_0(0)$ is finite and nonzero. It is easy to verify that the approximations (5.18) and (5.19) satisfy the time-independent Smoluchowski equation on the $O(\epsilon^{1/2})$ lengthscale near their respective turning points.

The appropriate (generic) approximation to the quasistationary density ρ_1 near the saddle is slightly more complicated; it is an error function approximation of the sort first used by Kramers [22]. We have [29]

$$\rho_1(x,y) \sim \text{const} \times \left\{ \epsilon^{-1/2} \int_0^\infty \exp\left[-(xp_x + p_x^2/4\lambda_x)/\epsilon \right] \, dp_x \right\} e^{-|\lambda_y|y^2/\epsilon} \tag{5.20}$$

$$= \text{const} \times \operatorname{erf}(\lambda_x^{1/2} x/\epsilon^{1/2}) e^{+\lambda_x x^2/\epsilon} e^{-|\lambda_y|y^2/\epsilon} \tag{5.21}$$

on the $O(\epsilon^{1/2})$ lengthscale. This 'boundary layer' approximation agrees with the inverted Gaussian approximation (5.19) in the far field, *i.e.*, as $x/\epsilon^{1/2} \to +\infty$. So under the same conditions, the tube approximation (5.16) will match to it.

We have now approximated $\rho_1(\boldsymbol{x})$ at all points $\boldsymbol{x}$ in the vicinity of the line segment joining S and the saddle. We must emphasize that the validity of this procedure depends on two assumptions:

- That the physical values of $W(\boldsymbol{x})$ and $K(\boldsymbol{x})$ at all points $\boldsymbol{x}$ along the axis arise from integration along the *on-axis* instanton trajectory extending from S to $\boldsymbol{x}$.

- That the WKB tube surrounding the axis is well behaved as the saddle is approached, so that the error function approximation to the quasistationary density is valid near the saddle. This requires that $w_2 \to 2|u_1(0)|$, and that k_0 tend to a finite, nonzero limit.

The first assumption breaks down when the MPEP has bifurcated, and we shall see that the second assumption breaks down at the onset of bifurcation. But if both assumptions hold, it is easy to compute the weak-noise asymptotics of the quasistationary eigenvalue λ_1 and its asymptotic reciprocal, the MFPT $\langle\tau\rangle$. The time-dependent equation $\dot\rho = -\mathcal{L}^*\rho$ may be written as

$$\dot\rho + \boldsymbol{\nabla}\cdot[-(\epsilon/2)\boldsymbol{\nabla}\rho + \rho\boldsymbol{u}] = 0. \tag{5.22}$$

Equation (5.22) is a continuity equation, and $\boldsymbol{j} = -(\epsilon/2)\boldsymbol{\nabla}\rho + \rho\boldsymbol{u}$ can be viewed as a probability current density. Since λ_1 is the decay rate of the eigenmode ρ_1, it may be computed as the rate at which probability is absorbed on the separatrix [22, 33]. Necessarily

$$\lambda_1 = \int_{-\infty}^{\infty}[-j_x(0,y)]\,dy \Bigg/ \int_0^{\infty}\int_{-\infty}^{\infty}\rho_1(x,y)\,dy\,dx \tag{5.23}$$

where $\boldsymbol{j} = (j_x, j_y)$ is computed from ρ_1. The numerator (an absorption rate) is computed from (5.21), and the normalization factor in the denominator from the Gaussian approximation (5.18). If the constant prefactors of these two approximations are chosen to ensure consistency with the intermediate WKB tube approximation (5.16), the quotient will acquire a factor $k_0(0)\exp[-w_0(0)/\epsilon]$, *i.e.*, $K(0,0)\exp[-W(0,0)/\epsilon]$.

This computation, if carried through, yields a so-called Eyring formula for the weak-noise asymptotics of the quasi-stationary eigenvalue, *i.e.*, the weak-noise asymptotics of the rate of noise-activated hopping [8, 15]:

$$\lambda_1(\epsilon) \sim K(0,0)\left[\frac{\sqrt{\lambda_x|\nu_x|}\sqrt{|\nu_y|/|\lambda_y|}}{\pi}\right]\exp[-\Delta W/\epsilon], \qquad \epsilon\to 0. \tag{5.24}$$

Here the presence of the 'frequency factor' $K(0,0)$ is attributable to the non-gradient dynamics; it will equal unity if the drift $\boldsymbol{u}$ is a gradient. The formula is otherwise familiar. Since $\langle\tau\rangle \sim \lambda_1^{-1}$ as $\epsilon\to 0$, this formula predicts a *pure Arrhenius* growth of the MFPT in the weak-noise limit. But as noted, this conclusion depends crucially on the validity of the Kramers-type error function approximation to the quasistationary density near the saddle. This approximation will prove not to be valid in double well models undergoing a bifurcation.

5.3 Indications of a Phase Transition

We shall now explain how the bifurcation transition displays characteristic features of a phase transition, such as power-law divergences governed by critical exponents. We begin by using the standard double well model (3.1), and the transport equations of the last section, to reveal the nature of the 'nascent cusp' singularity appearing at the saddle point, at criticality.

For the standard model, the stable point S is located at $x = x_s = 1$, and the coefficient functions (drift velocity derivatives) in the transport equations are of the form

$$v_0(x) = u_x(x,0) = x - x^3 \tag{5.25}$$

$$2v_2(x) = \partial^2 u_x/\partial y^2(x,0) = -2\alpha x \tag{5.26}$$

$$u_1(x) = \partial u_y/\partial y(x,0) = -\mu(1+x^2). \tag{5.27}$$

These may be substituted into the Riccati equation (5.8) for the transverse second derivative $w_2(x) = \partial^2 W/\partial y^2(x,0)$, and the equation numerically integrated. As noted, the appropriate initial condition is $w_2(x = x_s) = 2|u_1(x_s)|$, *i.e.*, $w_2(x = 1) = 4\mu$. Consider the case $\mu = 1$ (the subject of Fig. 2), in particular. One finds for all α in the range $0 < \alpha < 4$ that w_2 is positive on the line segment between $x = x_s$ and the saddle at $x = 0$. Since the WKB tube centered on the axis, which is formed by small transverse fluctuations about the MPEP, has variance $\sim \epsilon/w_2(x)$, this positivity implies that the tube is everywhere well-defined. One also finds that $w_2 \to 2$, *i.e.*, $w_2 \to 2|u_1(0)|$, as the saddle is approached. Moreover, by integrating the transport equation (5.13) one finds that $k_0(x) = K(x,0)$ tends to a finite, nonzero limit as $x \to 0$. As we explained above, these two conditions are precisely what is needed to ensure Arrhenius weak-noise asymptotics, with an MFPT prefactor proportional to $K(0,0)^{-1}$.

The bifurcation transition present in the $\mu = 1$ standard model at $\alpha = 4$ is reflected in the behavior of w_2 and k_0 as $x \to 0$. When $\alpha = \alpha_c = 4$, equations (5.8) and (5.13) can be solved exactly; one finds

$$w_2(x) = \partial^2 W/\partial y^2(x,0) = 4x^2, \tag{5.28}$$

$$k_0(x) = K(x,0) = 1/x. \tag{5.29}$$

We know from the Eyring formula that the activation rate prefactor, in the limit of weak noise, is proportional to $K(0,0)$. The fact that $k_0(x = 0) = K(0,0)$ is *infinite* here strongly suggests that at criticality the activation rate, *i.e.*, the rate at which the quasistationary density is absorbed on the separatrix, is *anomalously large*. Equivalently, it suggests that at criticality the weak-noise behavior of the MFPT (which is asymptotically equal to λ_1^{-1}) is *non-Arrhenius*, with a pre-exponential factor that tends to zero as $\epsilon \to 0$. There is an even stronger piece of evidence that this is the case. It is not difficult to show, by analysing the transport equation (5.13), that $K(0,0) = k_0(x = 0) \sim (\alpha_c - \alpha)^{-1/2}$ as $\alpha \to \alpha_c^-$. Figure 5 shows the result of a numerical computation when $\mu = 1$. *The activation rate prefactor diverges as* $\alpha \to \alpha_c^-$. Equivalently, the MFPT prefactor tends to zero. The natural deduction is that at criticality the activation rate prefactor, and its reciprocal the MFPT prefactor, become ϵ*-dependent*. This blends nicely with the behavior *above* the transition, since (as shown in Fig. 3) the exponential growth rate of the MFPT

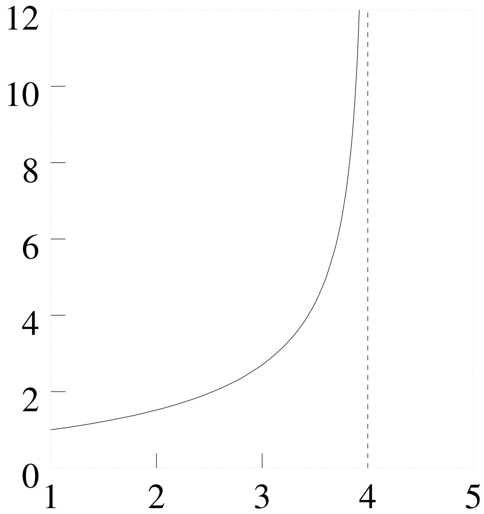

Figure 5: A plot of $K(0,0)$, to which the weak-noise activation rate prefactor is proportional, as a function of the off-axis softening parameter α. As shown, $K(0,0)$ diverges as $\alpha \to \alpha_c^-$. This is for the standard double well model (3.1), with $\mu = 1$, for which $\alpha_c = 4$.

(the action barrier ΔW, *i.e.*, $W(0,0)$) begins to decrease as α increases beyond α_c. An ϵ-dependent activation rate prefactor at criticality, containing a negative power of ϵ, would unify the exit behavior both below and above criticality.

We shall show in Section 8 that in critical double well models (*e.g.*, in the standard model with $\alpha = \alpha_c$), the weak-noise activation rate $\lambda_1 = \lambda_1(\epsilon)$ indeed has asymptotics

$$\lambda_1(\epsilon) \sim \text{const} \times \epsilon^{-s} \exp\left[-\Delta W/\epsilon\right], \qquad \epsilon \to 0, \tag{5.30}$$

where $s > 0$ is the singularity index mentioned in Section 4. The computation of the singularity index is nontrivial. Since the MFPT $\langle\tau\rangle$ satisfies $\langle\tau\rangle \sim \lambda_1^{-1}$, the ϵ^{-s} prefactor in λ_1 gives rise to an ϵ^s MFPT prefactor. In critical double well models, in the weak-noise limit the growth of the MFPT is *slower than pure exponential*.

At criticality, the action W as well as the prefactor K displays unusual behavior at the saddle. We shall see that the behavior of (5.28), *i.e.*, that w_2 tends to zero quadratically as $x \to 0$, is universal. Since the WKB tube has variance $\sim \epsilon/w_2(x)$, this implies that the tube *splays out* as the saddle is approached. To leading order, it splays out to infinite width. This is an indication that the transverse fluctuations around the MPEP, on the $O(\epsilon^{1/2})$ lengthscale, at criticality become very strong near the saddle. In fact, that $w_2(x=0) = 0$ causes some difficulty in the interpretation, near the saddle, of the WKB approximation to ρ_0 and ρ_1. One might expect that even though $w_2(x=0) = 0$, the quartic tube approximation

$$\rho_0(x,y) \sim k_0(x) \exp\left\{-\left[w_0(x) + w_2(x)\frac{y^2}{2!} + w_4(x)\frac{y^4}{4!}\right]/\epsilon\right\}, \qquad \epsilon \to 0 \tag{5.31}$$

would suffice for an understanding of the behavior of the WKB approximation near the saddle. If $w_2(x = 0)$ were zero but $w_4(x = 0)$ were finite and nonzero, transverse fluctuations around the saddle would be, by (5.31), of magnitude $O(\epsilon^{1/4})$ rather than $O(\epsilon^{1/2})$. However, explicit solution of eq. (5.15), the transport equation for $w_4 = \partial^4 W/\partial y^4 = \partial^3 p_y/\partial y^3$, shows that in the $\mu = 1$ standard model at $\alpha = \alpha_c = 4$,

$$w_4(x) \sim (4/5)x^{-4} + (16/5)x^{-2} + 8 + \cdots, \qquad x \to 0^+. \tag{5.32}$$

The fact that $w_4(x = 0)$ is *infinite*, coupled with the fact that $w_2(x = 0)$ is zero, suggests that at criticality, the transverse fluctuations near the saddle *have no natural scale*. In any event, at criticality the standard matched asymptotic approximations technique of the last section breaks down. We shall need to construct an approximation to the quasistationary density ρ_1 near the saddle which (i) is valid at criticality, and (ii) matches to the WKB approximation (5.31), despite its singular character.

Critical exponents, as we define them, describe the weak-noise behavior of a parametrized double well model with a singularity at some point $\boldsymbol{x}_0$, as $\boldsymbol{x} \to \boldsymbol{x}_0$ and as the parameters of the model tend to the values for which the singularity appears at $\boldsymbol{x}_0$. In particular, they characterize the behavior at and near the bifurcation transition, and at and near the saddle point, of the functions W and K appearing in the WKB approximation to ρ_0 and ρ_1. At criticality, the divergence rates of the WKB prefactor K as $x \to 0$ and $y \to 0$ supply two such exponents; the scaling form which we shall use to approximate W near a nascent cusp (which involves fractional powers) will supply others. There are also critical exponents describing what happens as one moves off criticality. As α is increased above α_c (*i.e.*, above 4, in the $\mu = 1$ standard model), the MPEP bifurcates. There is a critical exponent describing the separation rate of the two resulting MPEP's, as Fig. 3 makes clear. There is also a critical exponent describing the divergence rate of $K(0, 0)$ as $\alpha \to \alpha_c^-$, which as we have already noted equals $1/2$. The singularity index s can be regarded as a critical exponent too, though of a different kind; to compute it, one must go beyond the WKB approximation.

The 'nascent cusp,' as a singularity, is located in a space parametrized by x, y, and the parameter(s) of the drift field $\boldsymbol{u}$. But if one restricts oneself to a single double well model, the only parameters are x and y. In this case there is a natural analogy between the Lagrangian manifold $\mathcal{M}$ in phase space, formed by bicharacteristics emanating from $(S, \mathbf{0})$ and $(S', \mathbf{0})$, and a *thermodynamic surface*. The action W, as a function on the manifold, corresponds to a thermodynamic potential, in fact a Gibbs free energy. The equation $\boldsymbol{p} = \boldsymbol{p}(\boldsymbol{x}) = \partial W/\partial \boldsymbol{x}$ corresponds to a relation between conjugate state variables, such as pressure and volume. The singularities (points of non-differentiability) of the physical action $W(\boldsymbol{x})$ therefore correspond to *phase transitions*. The order of such a phase transition, in the traditional sense, is the lowest order of spatial derivative (of W) which fails to be continuous. In Section 8 we shall compute the order of the nascent cusp.

5.4 The Bifurcation Transition and Classical Mechanics

We now explain how the equations of Section 5.2 allow the bifurcation transition to be interpreted in terms of classical mechanics, and how one can predict whether or not any given double well model is at criticality. We begin by considering models in which the MPEP has already bifurcated,

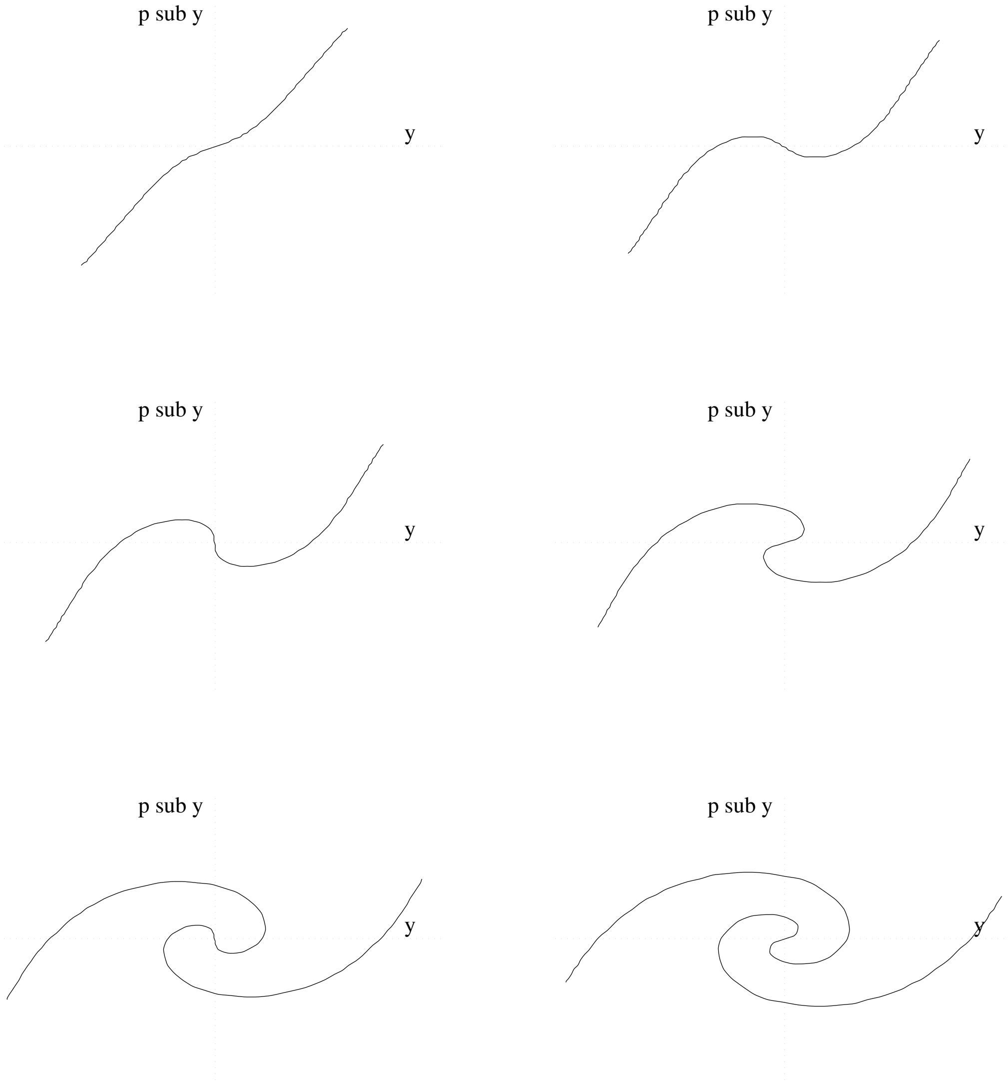

Figure 6: Cross-sections through the Lagrangian manifold $\mathcal{M}$, revealing the 'whorling' that takes place as one passes through any on-axis focus. These sketches show the map $y \mapsto p_y$ at successively decreasing values of x, as one moves from $S = (x_s, 0)$, past two foci $(x_f^{(1)}, 0)$ and $(x_f^{(2)}, 0)$, toward the saddle point $(0, 0)$. Shown are the cases (a) $x_s > x > x_f^{(1)}$ and x near x_s, (b) $x_s > x > x_f^{(1)}$ and x near $x_f^{(1)}$, (c) $x = x_f^{(1)}$, (d) $x_f^{(1)} > x > x_f^{(2)}$, (e) $x = x_f^{(2)}$, and (f) $x_f^{(2)} > x > 0$.

and the instanton trajectories emanating from the stable point S focus along the axis, as in Figs. 2(c) and 2(d).

Empirically, focusing occurs in models with a sufficiently large 'off-axis softening' parameter α. In such models, the action W in a mathematical sense becomes multivalued near a portion of the axis. Points $(x, 0)$ beyond the first focus are reached by multiple off-axis instanton trajectories emanating from the stable point S, and in general these trajectories will have different actions. They will also have different momenta $\boldsymbol{p} = \boldsymbol{\nabla} W$ at the time they reach $(x, 0)$.

This multivaluedness has a geometric interpretation, in terms of the shape of the two-dimensional Lagrangian manifold (in the four-dimensional phase space) formed by the bicharacteristics emanating from the point $(\boldsymbol{x}, \boldsymbol{p}) = (S, \boldsymbol{0})$. As x decreases from x_s toward zero, the map $y \mapsto p_y$ in the vicinity of $y = 0$ is at first single-valued; the value $p_y = 0$, and no other, corresponds to $y = 0$. Beyond the first focus $(x_f^{(1)}, 0)$, *i.e.*, when $x < x_f^{(1)}$, the map $y \mapsto p_y$ becomes three-valued. At the second focus $(x_f^{(2)}, 0)$ it becomes five-valued, etc. The generic evolution is shown in Figs. 6(a) through 6(f). Up to the first focus $y = y(p_y)$ near $p_y = 0$ may be modelled as a *linear* function; beyond the focus, as a *cubic*. Beyond the second focus the global description becomes more complicated, as is clear from the whorl in Fig. 6(f). A cubic approximation is still appropriate in the immediate vicinity of $(y, p_y) = (0, 0)$, however.

Since the locus of all points (y, p_y) at constant x is obtained by intersecting the Lagrangian manifold with the hyperplane $x = \text{const}$, the manifold itself becomes increasingly 'whorled' with each passage through a focus. The formation of convolutions in Lagrangian manifolds was first considered by Berry and Balazs [5] (in a time-dependent context), and the progression in Fig. 6 resembles the figures in their paper. Geometrically, the linear-to-cubic transition at each successive focus corresponds to the creation of a *fold* [12]. One can fit the shape of the manifold near the l'th focus, *i.e.*, near $(x, y, p_y) = (x_f^{(l)}, 0, 0)$, by the phenomenological formula

$$y = y(x, p_y) = -a_0^{(l)} p_y^3 - a_1^{(l)} \left(x - x_f^{(l)}\right) p_y \tag{5.33}$$

where $a_0^{(l)}$ and $a_1^{(l)}$ are certain positive constants. So each successive focus resembles a Ginzburg-Landau *second-order phase transition* (x corresponding to temperature, the focus location $x_f^{(l)}$ to a critical temperature, $-y$ to a magnetic field, and p_y to a magnetization). We shall say more about the 'equation of state' (5.33) (which we stress is *not* applicable near the 'nascent cusp' appearing at the saddle point of critical models) in Sections 7 and 8.

For on-axis (*i.e.*, $y = 0$, $p_y = 0$) trajectories, the derivative $w_2(x) = \partial p_y / \partial y(x, y = 0)$ satisfies the Riccati equation (5.14). So the appearance of a focus, and of multiple foci, can be investigated analytically. It is clear from Figs. 6(c) and 6(e) that passage through a focus is signalled by the tangent plane to the manifold (at $y = 0$, $p_y = 0$) 'turning vertical'; equivalently, by $\partial y / \partial p_y$ passing through zero, or its reciprocal w_2 (a negative magnetic susceptibility, in this context) passing through $-\infty$. To study this, recall that the Riccati equation

$$\dot{w}_2 = -w_2^2 - 2u_1 w_2 + 4v_0 v_2 \tag{5.34}$$

involves a derivative with respect to instanton transit time, and that the on-axis instanton trajectory (directed anti-parallel to the drift toward S) satisfies $\dot{x} = -v_0(x)$. Solutions w_2 can be regarded either as a function of t, for $-\infty < t < \infty$, or of x, for $x_s > x > 0$. We see from the form of the Riccati equation that w_2 can indeed be driven to $-\infty$ in finite time, *i.e.*, at some point $x = x_f > 0$ to the right of the saddle. In fact one sees, if t_f is the time when this occurs (the focus time), that as $t \to t_f^-$, *i.e.*, $x \to x_f^+$,

$$\begin{aligned} w_2(x) = \partial^2 W/\partial y^2(x,0) &\sim -(t_f - t)^{-1}, \\ &\sim \text{const} \times [-(x - x_f)^{-1}]. \end{aligned} \tag{5.35}$$

Here the constant multiplier equals $1/v_0(x_f)$, the reciprocal speed of the on-axis instanton trajectory when it passes through the focus $(x_f, 0)$. We note in passing that by the transport equation (5.13), this blowup will induce a blowup of the on-axis WKB prefactor k_0. One finds

$$\begin{aligned} k_0(x) = K(x,0) &\sim \text{const} \times (t_f - t)^{-1/2}, \\ &\sim \text{const} \times (x - x_f)^{-1/2}. \end{aligned} \tag{5.36}$$

Equations (5.35)–(5.36) contrast markedly with eqs. (5.28)–(5.29), which apply to the $\mu = 1$ standard model at criticality (where, in a formal sense, $x_f = 0$, since there is a *nascent* focus at the saddle). Equations (5.35)–(5.36) are not restricted to the standard model; they hold in greater generality. But they apply only when a bona fide focus is present at some $x_f > 0$ (*i.e.*, before the saddle is reached), and the MPEP has already bifurcated.

By examining the Riccati equation (5.34), we see that w_2 will be driven to $-\infty$, and a focus will be present, only if the inhomogeneous term $4v_0v_2$ on the right-hand side of (5.34) is sufficiently negative. (This is because $u_1 < 0$, by assumption.) But $2v_2 = \partial^2 u_x/\partial y^2(x, y = 0)$, which we are taking to be negative when $0 < x < x_s$, measures the extent to which the drift toward S softens as one moves off-axis. So our empirical observation is confirmed analytically: a sufficiently strong off-axis softening will create a focus, and a bifurcation of the MPEP!

It is best to think of $w_2 = \partial p_y/\partial y(y = 0)$ as a *slope*, as in Fig. 6. As such, it may rotate repeatedly through the point at infinity as t increases, *i.e.*, as x decreases. Each such rotation results in increased whorling of the Lagrangian manifold, and also corresponds to a passage through a focus. So by counting the number of singularities of the solution curve $w_2 = w_2(t)$, one may determine the number of foci present in any given double well model.

The standard model (3.1) will serve as an example. For the reasons discussed in Section 5.2 $w_2(t = -\infty)$, *i.e.*, $w_2(x = x_s)$, in the standard model always equals $2|u_1(x_s)| = 2|u_1(1)| = 4\mu$. Suppose that $\mu = 1$. We noted in the last section that if $0 < \alpha < 4$, w_2 is well-behaved and positive at all times t between $-\infty$ and ∞ inclusive, *i.e.*, at all x satisfying $0 \le x \le 1$. We also explained what happens at $\alpha = \alpha_c = 4$, when the nascent cusp appears at the saddle and the MPEP begins to bifurcate. At criticality, $w_2 \to 0$ as $t \to \infty$, *i.e.*, as $x \to 0^+$. If $4 < \alpha < 9$, w_2 is driven negative (as t increases toward ∞), and passes through $-\infty$ before returning (through $+\infty$) to finite, positive values. The change in behavior is shown in Fig. 7. When α is raised above α_c, we say that the graph of w_2 acquires *unit winding number*, since it winds once through the point at infinity. A second transition occurs at $\alpha = \alpha_c^{(2)} = 9$. If $9 < \alpha < 16$, w_2 passes through $-\infty$ *twice*, and its graph has winding number equal to 2. Except at the critical values $\alpha = \alpha_c^{(j)} = (j+1)^2$, w_2 in this model converges to the generic value $2|u_1(0)| = 2\,|\lambda_y| = 2$ as $t \to \infty$, *i.e.*, as $x \to 0^+$.

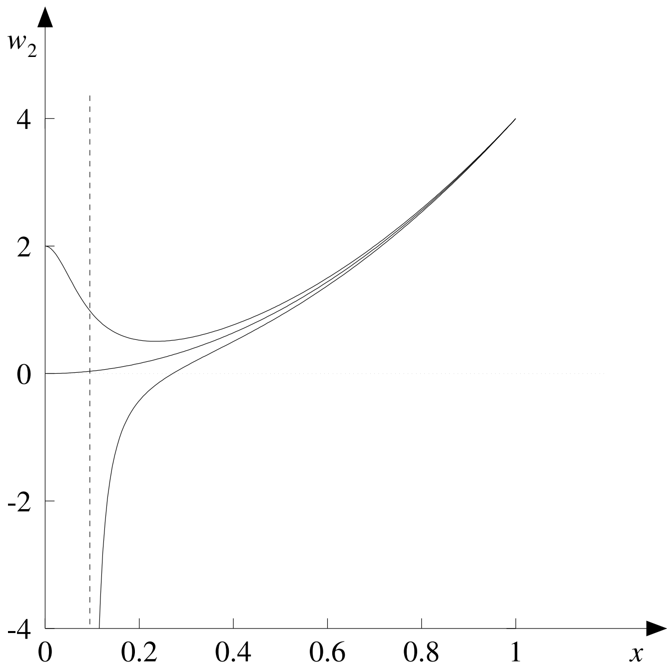

Figure 7: The transverse action derivative $w_2(x) \equiv \partial^2 W/\partial y^2(x,0)$ in the $\mu = 1$ variant of the standard double well model (3.1), in the vicinity of the bifurcation transition. The three curves, from top to bottom, obtain when $\alpha = 3.9$, 4.0 (the critical value α_c), and 4.1. When $\alpha < \alpha_c$, $w_2 \to 2|\lambda_y|$, *i.e.*, 2, as $x \to 0^+$. When $\alpha > \alpha_c$, w_2 is driven negative as x decreases, and passes through $-\infty$. The focal point $x = x_f$ where this occurs, in the model with $\alpha = 4.1$, is indicated by a dashed line.

Since each passage of w_2 through $-\infty$ gives rise to a focus, the sequence of near-axis instanton flow fields in the $\mu = 1$ standard model, as α is increased, displays an progressively larger number of foci. In fact the progression is precisely as displayed in Figs. 2(a) through 2(d). It is worth noting that in models with one or more foci, the WKB tube centered on the axis becomes ill-defined when w_2 goes negative, which takes place at a location on the axis somewhat before the first focus is reached. (See Fig. 7.)

In Section 8 we shall determine exactly what happens at the bifurcation transition of any singly parametrized symmetric double well model. But we can now pose the question: What, physically, causes the above values for α to be critical? If in general the odd function $2v_2(x) = \partial^2 u_x/\partial y^2(x, 0)$ is negative between $x = 0$ and $x = x_s$ and is proportional to a parameter α, is there a classical mechanical technique of predicting the values of α at which the on-axis instanton trajectory will bifurcate? The answer to this question is 'yes.' Our technique relies on a *linear stability analysis* of the on-axis instanton trajectory, and identifies the critical values of α as the values for which a transverse *soft mode* is present in the zero-energy Hamiltonian dynamics. This is reminiscent of Langer's analysis of metastability in one-dimensional models [24, 36]. But because we shall consider transverse, rather than longitudinal, fluctuations around the instanton trajectory, our stability analysis will be considerably simplified.

Let $\boldsymbol{x} = \boldsymbol{x}^*(t) = (x^*(t), 0)$ be the on-axis instanton trajectory, where $x = x^*(t)$ is the solution of $\dot{x} = -v_0(x)$. Near-axis instanton trajectories, *i.e.*, near-axis zero-energy classical trajectories emanating from S, may to leading order be written as

$$\boldsymbol{x} = \boldsymbol{x}(t) = (x(t), y(t)) \sim (x^*(t), 0) + \delta\,(0, Y(t)) \tag{5.37}$$

where $\delta \ll 1$, and where $Y = Y(t)$ is some model-dependent function satisfying $Y(t = -\infty) = 0$. $Y(t)$, $-\infty < t < \infty$, is a *normalized transverse deviation*. Similarly, near-axis trajectories have momenta

$$\boldsymbol{p} = \boldsymbol{p}(t) = (p_x(t), p_y(t)) \sim (p_x^*(t), 0) + \delta\,(0, P_y(t)) \tag{5.38}$$

for some unknown function $P_y = P_y(t)$ satisfying $P_y(t = -\infty) = 0$. Here $\boldsymbol{p} = \boldsymbol{p}^*(t) = (p_x^*(t), 0)$ is the momentum of the on-axis instanton trajectory at instanton transit time t. We noted before eq. (5.12) that as a function of x, p_x^* equals $-2v_0$. So $p_x^*(t)$ equals $-2v_0\,(x^*(t))$. We necessarily have

$$w_2(t) = P_y(t)/Y(t), \tag{5.39}$$

on account of w_2 equalling $\partial p_y/\partial y(y = 0)$.

Substituting equations (5.37) and (5.38) into the Hamilton equations derived from the Wentzell-Freidlin Hamiltonian H, and separating terms proportional to δ, yields the pair of equations

$$\dot{Y} = \frac{\partial^2 H}{\partial p_y \partial y}\,(\boldsymbol{x}^*(t), \boldsymbol{p}^*(t))\,Y + \frac{\partial^2 H}{\partial p_y^2}\,(\boldsymbol{x}^*(t), \boldsymbol{p}^*(t))\,P_y \tag{5.40}$$

$$\dot{P}_y = -\frac{\partial^2 H}{\partial y^2}\,(\boldsymbol{x}^*(t), \boldsymbol{p}^*(t))\,Y - \frac{\partial^2 H}{\partial p_y \partial y}\,(\boldsymbol{x}^*(t), \boldsymbol{p}^*(t))\,P_y. \tag{5.41}$$

Due to the special form of the Hamiltonian (5.2), and the expansions (5.9), this pair becomes

$$\dot{Y} = u_1\left(x^*(t)\right) Y + P_y \tag{5.42}$$

$$\dot{P}_y = 4v_0\left(x^*(t)\right) v_2\left(x^*(t)\right) Y - u_1\left(x^*(t)\right) P_y. \tag{5.43}$$

So $w_2 = w_2(t)$ can be represented as the quotient of two functions of instanton transit time, which satisfy a pair of coupled *linear* differential equations. We note in passing that an analogous representation is possible for solutions $\boldsymbol{Z} = \boldsymbol{Z}(t)$ of the matrix Riccati equation (5.8). The existence of such quotient representations is well known in the theory of Riccati equations, and has a geometric interpretation [37].

Equations (5.40)–(5.41), and (5.42)–(5.43), may be viewed as Hamilton's equations for the *effective* (*i.e.*, time-dependent) *transverse Wentzell-Freidlin Hamiltonian*

$$\begin{aligned} H_{\text{eff}}(Y, P_y, t) &= \frac{1}{2}\frac{\partial^2 H}{\partial p_y^2} P_y^2 + \frac{\partial^2 H}{\partial p_y \partial y} Y P_y + \frac{1}{2}\frac{\partial^2 H}{\partial y^2} Y^2 \\ &= P_y^2/2 + u_1\left(x^*(t)\right) Y P_y - 2v_0\left(x^*(t)\right) v_2\left(x^*(t)\right) Y^2. \end{aligned} \tag{5.44}$$

To save space we have suppressed the arguments $(\boldsymbol{x}^*(t), \boldsymbol{p}^*(t))$ of the partial derivatives. This quadratic Hamiltonian governs the small transverse fluctuations about the on-axis instanton trajectory. Its Legendre transform $\dot{Y} P_y - H_{\text{eff}}$, namely

$$\begin{aligned} L_{\text{eff}}(Y, \dot{Y}, t) &= \frac{1}{2}\frac{\partial^2 L}{\partial \dot{y}^2} \dot{Y}^2 + \frac{\partial^2 L}{\partial \dot{y} \partial y} Y \dot{Y} + \frac{1}{2}\frac{\partial^2 L}{\partial y^2} Y^2 \\ &= |\dot{Y} - u_1\left(x^*(t)\right) Y|^2/2 + 2v_0\left(x^*(t)\right) v_2\left(x^*(t)\right) Y^2, \end{aligned} \tag{5.45}$$

is an *effective transverse Onsager-Machlup Lagrangian*. Here L is the Onsager-Machlup Lagrangian (5.3), and we have suppressed the arguments $(\boldsymbol{x}^*(t), \dot{\boldsymbol{x}}^*(t))$ of the partial derivatives. The corresponding Euler-Lagrange equation for the normalized transverse deviation Y, *i.e.*,

$$\ddot{Y} + \left\{ -\frac{d}{dt}\left[u_1\left(x^*(t)\right)\right] - u_1^2\left(x^*(t)\right) - 4v_0\left(x^*(t)\right) v_2\left(x^*(t)\right) \right\} Y = 0, \tag{5.46}$$

is called a (transverse) *Jacobi equation* [36]. It may be written as an equation for $Y = Y(x)$, $0 \le x \le x_s$, by changing the independent variable from t to x. One gets

$$\mathcal{J} Y \equiv \frac{d}{dx}\left[v_0(x)\frac{dY}{dx}\right] + \left[u_1'(x) - \frac{u_1^2(x)}{v_0(x)} - 4v_2(x)\right] Y = 0 \tag{5.47}$$

together with the boundary condition $Y(x = x_s) = 0$. This Jacobi equation, which is in Sturm-Liouville form, governs the behavior of the instanton trajectories near the on-axis trajectory (the MPEP, if it has not bifurcated). So it is responsible for the various behaviors shown in Fig. 2. It is clear from our derivation that the Jacobi operator $\mathcal{J}$, considered as a quadratic form, defines the transverse *second variation* of the Onsager-Machlup action functional about the on-axis trajectory.

Foci are by definition the points $(x, 0)$ where the off-axis instanton trajectories converge (to leading order). Equivalently, $(x, 0)$ is a focus only if $Y(x) = 0$. But since $w_2 = P_y/Y$, this implies that (unless $P_y(x) = 0$ also) $w_2(x)$ is infinite. This is precisely the necessary condition for a focus that we derived earlier. If Y passes through zero more than once, then w_2 will pass through the point at infinity more than once. This is the mechanism by which, *e.g.*, the two-focus flow field of Fig. 2(d) engenders the increasingly 'whorled' Lagrangian manifold of Figs. 6(a) through 6(f).

We can now give a simple criterion for determining whether or not a given double well model is at criticality. Suppose that the most probable escape path (MPEP) extends along the axis from S to the saddle, so that the symmetry is as yet unbroken. We know by the discussion in Section 4 that criticality is signalled by the appearance of a nascent cusp at the saddle. The nascent cusp itself is not a focus, as Fig. 2(b) makes clear. But if the off-axis softening is increased, the nascent cusp becomes a genuine cusp (*i.e.*, focus); it moves inward along the axis from the saddle toward S. This picture is consistent with the interpretation of the near-axis instanton flow field in terms of the function $Y(x)$, $0 \le x \le x_s$, only if the nascent cusp, like a conventional on-axis focus, is a zero of Y.

So the signal for criticality is Y equalling zero at $x = 0$. We can rephrase this as follows. *Critical double well models are those models with unbroken symmetry for which the Jacobi equation $\mathcal{J}Y = 0$ for the transverse deviation function Y, equipped with boundary condition $Y(x = x_s) = 0$ and also with $Y(x = 0) = 0$, has a nontrivial (*i.e.*, nonzero) solution.* The nonzero solution $Y = Y_1(x)$, $0 \le x \le x_s$, when it exists, can be interpreted as a transverse soft mode of the zero-energy Hamiltonian dynamics. If the off-axis softening is increased, the on-axis MPEP will bifurcate. Just beyond criticality, there will be two symmetrically placed off-axis MPEP's from S to the saddle. They will be of the form $(x^*(t), \pm\delta Y_1(x^*(t)))$, for some small δ. This 'motion in the direction of a soft mode' is a standard bifurcation effect. *At criticality, the transverse soft mode Y_1 describes the way in which the two MPEP's separate.*

Suppose that the double well model is parametrized by an off-axis softening parameter α, *i.e.*, that $v_2 = \alpha\hat{v}_2$ for some odd function $\hat{v}_2$, and that v_0 and u_1 are independent of α. Then by rewriting the Jacobi equation, one sees that the model will be at a bifurcation point if and only if the Sturm-Liouville equation

$$\hat{\mathcal{J}}Y \equiv \frac{1}{4\hat{v}_2(x)}\frac{d}{dx}\left[v_0(x)\frac{dY}{dx}\right] + \frac{1}{4\hat{v}_2(x)}\left[u_1'(x) - \frac{u_1^2(x)}{v_0(x)}\right]Y = \alpha Y, \tag{5.48}$$

equipped with Dirichlet boundary conditions $Y(x = 0) = Y(x = x_s) = 0$, has a nonzero solution. The Sturm-Liouville operator $\hat{\mathcal{J}}$ may be called a *normalized Jacobi operator*.

We see that the set of critical values of α is precisely the *spectrum* of the normalized Jacobi operator! Only the first critical value (*i.e.*, lowest eigenvalue) $\alpha_c = \alpha_c^{(1)}$ will yield an actual bifurcation of the MPEP. To each higher critical value $\alpha_c^{(j)}$, $j = 2, 3, \ldots$, there corresponds a transverse eigenmode Y_j. But after the first bifurcation, the on-axis instanton trajectory is no longer the physical MPEP. The higher eigenmodes Y_j, $j = 2, 3, \ldots$, which are oscillatory, govern the further bifurcations of the on-axis instanton trajectory rather than the further bifurcations (if any) of the physical MPEP's, which have already moved off-axis.

The case of the standard model (3.1) is instructive. Substituting from (5.25)–(5.27) one finds as normalized Jacobi operator

$$\hat{\mathcal{J}} = -\frac{1}{4x}\frac{d}{dx}(x - x^3)\frac{d}{dx} + \left[\frac{\mu}{2} + \frac{\mu^2(1+x^2)^2}{4x^2(1-x^2)}\right]. \tag{5.49}$$

It is easily verified that on the interval from $x = 0$ to $x = x_s = 1$, this operator (when equipped with Dirichlet boundary conditions) has spectrum

$$\alpha_c^{(j)} = j^2 + (3\mu - 1)j + (2\mu^2 - \mu), \qquad j = 1, 2, 3, \ldots \tag{5.50}$$

So in the standard model, the bifurcation of the physical MPEP occurs at $\alpha_c = \alpha_c^{(1)} = 2\mu(\mu + 1)$. Also, the standard model with $\mu = 1$ has $\alpha_c^{(j)} = (j+1)^2$, so the on-axis instanton trajectory bifurcates at $\alpha = 4, 9, 16, \ldots$ We have several times mentioned this curious progression of squares. The eigenfunctions Y_j corresponding to the eigenvalues $\alpha_c^{(j)}$, *i.e.*, the transverse soft modes appearing at $\alpha = \alpha_c^{(j)}$, turn out to be of the form

$$Y_j(x) = (x - x^3)^\mu q_j(x), \tag{5.51}$$

where q_j is an even polynomial of degree $2j - 2$. Substituting into the transverse Hamilton equation (5.42) yields the analogous transverse momentum deviations $(P_y)_j$. One gets

$$\begin{aligned}(P_y)_j(x) &= -v_0(x)Y_j'(x) - u_1(x)Y_j(x) \\ &= (x - x^3)^\mu \left[4\mu x^2 q_j(x) - (x - x^3)q_j'(x)\right],\end{aligned} \tag{5.52}$$

so that in the standard model at $\alpha = \alpha_c^{(j)}$,

$$w_2(x) = (P_y)_j(x)/Y_j(x) = 4\mu x^2 - (x - x^3)q_j'(x)/q_j(x). \tag{5.53}$$

If $j = 1$ then q_j reduces to a constant, and the second term is absent. So at the physical bifurcation point [*i.e.*, at $\alpha = \alpha_c = 2\mu(\mu + 1)$], $w_2(x)$ equals $4\mu x^2$. Moreover, $Y_1(x) = (x - x^3)^\mu$. This transverse soft mode is seen clearly in Figs. 2(c) and 2(d), which show the behavior of the $\mu = 1$ standard model beyond the bifurcation point. In those figures the off-axis MPEP's are roughly proportional to $\pm(x - x^3)$, *i.e.*, to $\pm Y_1$. *As α is increased above α_c, the MPEP's move in the direction of the transverse soft mode.*

Recall that the profile of the WKB tube of probability density centered on the x-axis is asymptotically Gaussian, and that at specified x this Gaussian has variance $\sim \epsilon/w_2(x)$. But in the standard model, at the first (and only physical) critical value $\alpha = \alpha_c = 2\mu(\mu + 1)$, $w_2(x)$ equals $4\mu x^2$. That $w_2 \to 0$ as $x \to 0$ implies that at criticality, the WKB tube *splays out* as the saddle is approached. We have already seen the $\mu = 1$ case of this in Section 5.3. The splayout is what one would expect from our picture of the bifurcation of the MPEP, which begins at the saddle, as a phase transition. It simply says that on the $O(\epsilon^{1/2})$ transverse lengthscale, the Gaussian fluctuations about the MPEP grow without bound as the nascent cusp is approached.

It is easy to see that this behavior is *universal*: it occurs in any critical double well model with a bifurcating MPEP. If the (diagonal) linearization of the drift field $\boldsymbol{u}$ at the saddle has eigenvalues (λ_x, λ_y), and μ is defined as usual to equal $|\lambda_y|/\lambda_x$, then examination of the Jacobi equation shows that the soft mode Y_1 has asymptotics $Y_1(x) \sim Cx^\mu$, $x \to 0^+$, for some nonzero constant C. We have mentioned this 'approach path' property elsewhere [28]. Also, examination of the Hamilton equation for $(P_y)_1$ shows that $(P_y)_1(x) \sim C'x^{\mu+2}$ for some nonzero C'. So at criticality, the quotient $w_2(x)$ satisfies (for any μ)

$$w_2(x) = \partial^2 W/\partial y^2(x, 0) \sim \text{const} \times x^2, \tag{5.54}$$

as $x \to 0^+$, and the tube splayout always occurs. Incidentally, it follows by integrating the transport equation (5.13) that

$$k_0(x) = K(x, 0) \sim \text{const} \times x^{-\mu}, \tag{5.55}$$

as $x \to 0^+$. Equations (5.54)–(5.55) summarize the universal behavior of the WKB tube near the saddle, in any critical double well model. They are the extension to arbitrary critical models of eqs. (5.28)–(5.29), which applied only to the critical variant ($\alpha = \alpha_c = 4$) of the $\mu = 1$ standard model.

We stressed in Section 5.2 that a Kramers-type error function approximation to the quasi-stationary density ρ_1 near the saddle is appropriate only if $w_2 \to 2|u_1(0)|$ as the saddle is approached. At criticality, since $w_2 \to 0$ instead, in order to apply the method of matched asymptotic approximations we shall need to construct a different boundary layer approximation. This will give rise to the universal non-Arrhenius MFPT asymptotics for models at criticality.

6 Maslov-WKB Asymptotics

By building on the previous sections, we can analyse the weak-noise behavior of double-well models with singularities. We have seen that singularities may appear in the WKB approximation $K(\boldsymbol{x}) \exp[-W(\boldsymbol{x})/\epsilon]$ for the stationary density ρ_0 and quasistationary density ρ_1. The possible singular behaviors are summed up in eqs. (5.35)–(5.36), which apply to models in which the MPEP has already bifurcated, and eqs. (5.54)–(5.55), which apply to models which are critical in the sense of bifurcations. Models in which the MPEP has already bifurcated have the property that the instanton trajectories emerging from S focus at a point $(x_f, 0)$ on the axis, with $x_f > 0$. The prefactor K of the WKB approximation will diverge there. In critical models, there is no actual on-axis focusing. But the prefactor will nonetheless diverge at the saddle point $(0, 0)$.

There is a standard procedure for extending the WKB approximation to such singular points, by 'glueing in' auxiliary, non-WKB approximations. It originated with the work of Keller and Rubinow on short-wave asymptotics [20], and has been most extensively developed by Maslov [32]. For a mathematically rigorous treatment, see Duistermaat [10]. See also Eckmann and Sénéor [13], for a partly pedagogical one-dimensional treatment. The procedure may be applied to the (formal) asymptotic solutions of any partial differential equation of the form $H(\boldsymbol{x}, -\epsilon\boldsymbol{\nabla})\rho = 0$, where H is a specified Hamiltonian. Here we discuss its application to the Smoluchowski equation, in arbitrary dimensionality n.

We know from Section 5.1 that mathematically, the WKB approximation to ρ_0 and ρ_1 is determined by (i) a Lagrangian manifold $\mathcal{M}$ in the $2n$-dimensional phase space, formed by the bicharacteristics emanating from $(S, \mathbf{0})$ and $(S', \mathbf{0})$, and (ii) functions W and K defined on this manifold, and computable by integration along the bicharacteristics. Of the points $\mathfrak{p}^{(i)} = (\boldsymbol{x}, \boldsymbol{p}^{(i)})$ 'over' any point $\boldsymbol{x}$, only the one with least action is physical. The values there of W and K are the values $W(\boldsymbol{x})$ and $K(\boldsymbol{x})$ appearing in the WKB approximation.

This geometric interpretation motivates the introduction of a new, 'diffraction integral' way of formulating the WKB approximation. At any point $\mathfrak{p} = (\boldsymbol{x}, \boldsymbol{p})$ on the Lagrangian manifold $\mathcal{M}$, we have

$$W(\mathfrak{p}) = \int \boldsymbol{p} \cdot d\boldsymbol{x}, \tag{6.1}$$

the line integral being taken along the bicharacteristic terminating at $\mathfrak{p}$. We can define a Legendre transform $\widetilde{W}$, satisfying $\widetilde{W} = \boldsymbol{x} \cdot \boldsymbol{p} - W$, by

$$\widetilde{W}(\mathfrak{p}) = \int \boldsymbol{x} \cdot d\boldsymbol{p}. \tag{6.2}$$

It is natural to think of $\widetilde{W}$ as a function of momentum $\boldsymbol{p}$, by projecting 'sideways' onto momentum space. Of course $\widetilde{W}(\boldsymbol{p})$ is potentially multivalued, like $W(\boldsymbol{x})$. For W, it is the *least* of the possible values that is physical; for $\widetilde{W}$, it is the *most*. But if one ignores the multivaluedness of $\widetilde{W}(\boldsymbol{p})$, one can write

$$K(\boldsymbol{x}) \exp[-W(\boldsymbol{x})/\epsilon] \sim \epsilon^{-n/2} \int \cdots \int \widetilde{K}(\boldsymbol{p}) \exp\left\{ \left[-\boldsymbol{x} \cdot \boldsymbol{p} + \widetilde{W}(\boldsymbol{p}) \right] / \epsilon \right\} dp_1 \cdots dp_n, \tag{6.3}$$

where $K(\boldsymbol{x})$ and $\widetilde{K}(\boldsymbol{p})$ are related by

$$K(\boldsymbol{x}) \propto \widetilde{K}(\boldsymbol{p}) \Bigg/ \sqrt{\det\left[-\frac{\partial^2 \widetilde{W}}{\partial p_i \partial p_j}(\boldsymbol{p}) \right]} = \widetilde{K}(\boldsymbol{p}) \sqrt{\det\left[-\frac{\partial^2 W}{\partial x_i \partial x_j}(\boldsymbol{x}) \right]}, \tag{6.4}$$

the correspondence between $\boldsymbol{p}$ and $\boldsymbol{x}$ being given by $\boldsymbol{p}(\boldsymbol{x}) = \partial W/\partial \boldsymbol{x}$, or $\boldsymbol{x}(\boldsymbol{p}) = \partial \widetilde{W}/\partial \boldsymbol{p}$. The asymptotic equality in (6.3), as $\epsilon \to 0$, is justified by the method of steepest descent. (It may be necessary to cut off the integral at large momentum to ensure convergence.) The method of steepest descent automatically picks out the point $\mathfrak{p} = (\boldsymbol{x}, \boldsymbol{p})$ 'over' $\boldsymbol{x}$ with the least action $W(\mathfrak{p})$. We shall call (6.3) a *diffraction integral representation*, since (if ϵ is pure imaginary) it resembles the diffraction integrals used in physical optics [4].

We have assumed that the Hessian matrix $\partial^2 W/\partial x_i \partial x_j = \partial p_i/\partial x_j$, whose inverse is the matrix $\partial^2 \widetilde{W}/\partial p_i \partial p_j = \partial x_i/\partial p_j$, is negative definite. Actually it is often possible to make sense of the above formulæ even when this is not the case, by analytic continuation. It is also possible to avoid the problem of positive eigenvalues by taking the Legendre transform with respect to a partial (incomplete) set of variables. In $n = 2$ dimensions, this means with respect to a single variable only. For example, one could use the alternative integral representation

$$K(x, y) \exp[-W(x, y)/\epsilon] \sim \epsilon^{-1/2} \int \widetilde{K}^{(x)}(p_x, y) \exp\left\{ \left[-x p_x + \widetilde{W}^{(x)}(p_x, y) \right] / \epsilon \right\} dp_x, \tag{6.5}$$

where $\widetilde{W}^{(x)} = xp_x - W$ is regarded as a function of p_x and y, and K and $\widetilde{K}^{(x)}$ are related by

$$K(x,y) \propto \widetilde{K}^{(x)}(p_x,y) \Bigg/ \sqrt{-\frac{\partial^2 \widetilde{W}^{(x)}}{\partial p_x^2}(p_x,y)} = \widetilde{K}^{(x)}(p_x,y)\sqrt{-\frac{\partial^2 W}{\partial x^2}(x,y)}. \tag{6.6}$$

Here the correspondence between (x,y) and (p_x,y) is given by $p_x(x,y) = \partial W/\partial x$, or equivalently $x(p_x,y) = \partial \widetilde{W}^{(x)}/\partial p_x$.

It is clear that the transformed prefactor $\widetilde{K}$ (resp. $\widetilde{K}^{(x)}$, etc.) in these integral representations, like W, K, and $\widetilde{W}$ (resp. $\widetilde{W}^{(x)}$, etc.), can be thought of as a function on the Lagrangian manifold $\mathcal{M}$. Also, the momentum integration can be viewed as an integration over $\mathcal{M}$. So introducing integral representations of this sort is really a way of replacing the position-space WKB approximation $K(\boldsymbol{x})\exp[-W(\boldsymbol{x})]$ to $\rho(\boldsymbol{x})$ by a smeared-out equivalent one, or ones, involving integration over the manifold. As derived, these 'momentum space' approximations are accurate only to leading order as $\epsilon \to 0$, since subdominant terms in ϵ arising from the method of steepest descent have been neglected. But such terms could be incorporated, if desired, by adding ϵ-dependent corrections to the transformed prefactor.

If the new formulations of the WKB approximation are equivalent to the old, why have we introduced them? The reason is that the equivalence holds only at points $\boldsymbol{x}$ at which K is finite. At singularities of K, the new formulations provide a means of computing the true $\epsilon \to 0$ asymptotics of ρ. Moreover, they reveal how at least some singularities of K can be explained as *artifacts*, arising from the way in which K is computed from $\widetilde{K}$. It follows from (6.4) that if the determinant of the Hessian matrix $\partial^2 W/\partial x_i \partial x_j = \partial p_i/\partial x_j$ diverges at some point $\boldsymbol{x}$, then $\boldsymbol{x}$ will be a singularity of K whenever $\widetilde{K}$ is nonzero at the corresponding momentum $\boldsymbol{p} = \boldsymbol{p}(\boldsymbol{x})$. In other words, singularities of K may be more apparent than real: they can arise from points $(\boldsymbol{x},\boldsymbol{p})$ on the manifold where $\widetilde{K}$ does not actually diverge. A similar effect can arise from the representation (6.5), or from any other diffraction integral representation.

The matrix $\partial p_i/\partial x_j$ is a matrix of partial slopes, which specifies (to first order) the shape of the manifold in the vicinity of the point $(\boldsymbol{x},\boldsymbol{p}) = (\boldsymbol{x},\boldsymbol{p}(\boldsymbol{x}))$. Its determinant becomes infinite only when at least one of its elements is infinite. Such a blowup occurs only at locations on the manifold where the (n-dimensional) tangent hyperplane to the manifold 'turns vertical,' *i.e.*, points along a momentum direction in the $2n$-dimensional phase space. This is precisely the behavior one sees at a *fold*, as in Figs. 2(c), 2(d), and 6. *The folding over of the Lagrangian manifold can create singularities of K.* This is clearly the cause of the singular behavior of K at the on-axis foci occurring in models with a bifurcated MPEP, though *not* of the singular behavior at the nascent cusp occurring at criticality (which cannot be transformed away).

Whether or not a singularity of the prefactor K occurring at some point $\boldsymbol{x} = \boldsymbol{x}^*$ is an artifact of this sort, by employing an appropriate diffraction integral representation one may compute the true weak-noise asymptotics of $\rho(\boldsymbol{x}^*)$. One usually finds leading-order behavior of the form const $\times\, \epsilon^{-\sigma} \exp[-W(\boldsymbol{x}^*/\epsilon)]$, where σ is by definition the singularity index of $\boldsymbol{x}^*$. In fact the ϵ-dependent prefactor $\epsilon^{-\sigma}$ should appear at all points $\boldsymbol{x}$ within some ϵ-dependent distance of $\boldsymbol{x}^*$, which shrinks to zero as $\epsilon \to 0$. Within this local region an asymptotically exact formula for ρ, derived from the integral representation, will be uniformly valid. This asymptotic approximation (non-WKB, at least in the traditional sense) will match in the far field to the WKB approximation $K(\boldsymbol{x})\exp[-W(\boldsymbol{x})/\epsilon]$.

In Sections 7 and 8 we shall see how this 'glueing in' procedure works, both in models with a bifurcated MPEP and in models at criticality. For the moment we note only that the construction of a local approximation to ρ, near the singular point $\boldsymbol{x}^*$, depends crucially on the determination of the behavior of $\widetilde{W}$ and $\widetilde{K}$ near the corresponding point $\boldsymbol{p}^*$ in momentum space. The case when $\widetilde{K}$ is well-behaved ('slowly varying') in a neighborhood of $\boldsymbol{p}^*$, and the singularity at $\boldsymbol{x} = \boldsymbol{x}^*$ is an artifact, is the simplest. Suppose that $\widetilde{W} = \widetilde{W}(\boldsymbol{p})$ can be expanded in a power series around $\boldsymbol{p} = \boldsymbol{p}^*$. The matrix $\partial^2\widetilde{W}/\partial p_i \partial p_j$ must have a zero eigenvalue at $\boldsymbol{p} = \boldsymbol{p}^*$, since otherwise the determinant of its inverse $\partial^2 W/\partial x_i \partial x_j$ would not tend to infinity as $\boldsymbol{x} \to \boldsymbol{x}^*$, the Lagrangian manifold would not turn vertical there, and the singularity in K would not appear. The term *catastrophe* is used to describe what happens to the manifold at $\boldsymbol{x} = \boldsymbol{x}^*$. It is a standard result, due largely to Arnol'd [2], that if the manifold is smooth near $(\boldsymbol{x}^*, \boldsymbol{p}^*)$, the catastrophic behavior at $\boldsymbol{x} = \boldsymbol{x}^*$ can be captured by approximating $\widetilde{W} = \widetilde{W}(\boldsymbol{p})$ by one of a handful of polynomial functions. These are the 'structurally stable' *elementary catastrophes*.

A single example, illustrating the similarity to the Ginzburg-Landau theory of phase transitions, will suffice. In n dimensions, suppose that a singularity at $\boldsymbol{x} = \boldsymbol{x}^*$ arises as an artifact in the above sense, and that $\boldsymbol{p}^* = \partial W/\partial \boldsymbol{x}(\boldsymbol{x}^*)$. In appropriate (linearly transformed) coordinates, write

$$\boldsymbol{z} = (z_1, \ldots, z_n) = \boldsymbol{x} - \boldsymbol{x}^* \tag{6.7}$$

$$\boldsymbol{g} = (g_1, \ldots, g_n) = \boldsymbol{p} - \boldsymbol{p}^*. \tag{6.8}$$

A particularly common sort of catastrophe (a 'cuspoid') would be described locally by a single-variable Legendre transform of the form

$$\widetilde{W}^{(z_n)}(z_1, \ldots, z_{n-1}, g_n) = -\frac{a_0 g_n^{n+2}}{n+2} - \frac{a_1 z_1 g_n^n}{n} - \cdots - \frac{a_{n-1} z_{n-1} g_n^2}{2} + R(z_1, \ldots, z_{n-1}) \tag{6.9}$$

where $a_0, \ldots, a_{n-1}$ are constants, and $R(z_1, \ldots, z_{n-1})$ is a quadratic polynomial. Since $z_n = \partial\widetilde{W}^{(z_n)}/\partial g_n$, this expression implies

$$z_n = z_n(z_1, \ldots, z_{n-1}, g_n) = -a_0 g_n^{n+1} - a_1 z_1 g_n^{n-1} - \cdots - a_{n-1} z_{n-1} g_n. \tag{6.10}$$

The presence of a catastrophe at $(z_1, \ldots, z_{n-1}, g_n) = (0, \ldots, 0, 0)$ is signalled by the fact that $\partial^2\widetilde{W}/\partial g_n^2 = \partial z_n/\partial g_n$ equals zero there.

We have already seen the $n = 2$ version of eq. (6.10) in Section 5.4, as a phenomenological description of the shape of the manifold $\mathcal{M}$ near an on-axis focus. Recall that we interpreted eq. (5.33), which is the $n = 2$ version, in thermodynamic terms: as the equation of state of a substance undergoing a Ginzburg-Landau second-order phase transition. (*E.g.*, z_1 is $T - T_c$, z_2 is a negative magnetic field, and g_2 is magnetization.) Equation (6.10) is in fact a *normal form* for the shape of a Lagrangian manifold near a cuspoid singularity. When $n = 2$, the cuspoid is a *cusp*. If $n = 1$, only the first term on the right-hand side of (6.10) is present, and the cuspoid reduces to a quadratic *fold*.

In general, to each possible polynomial expression (normal form) for the Legendre-transformed action, there corresponds a non-WKB approximation to ρ in a local region near $\boldsymbol{x} = \boldsymbol{x}^*$, computed from the appropriate diffraction integral. These integrals serve to define the *canonical diffraction functions* first explored by Maslov. The canonical diffraction functions include the classical Airy and Pearcey functions, which arise from folds and cusps respectively [4]. We shall study the cusp case further in the next section, as a warmup for the study of the nascent cusp appearing at criticality. The normal form for the action near a nascent cusp will turn out to be nonpolynomial, but the Maslov-WKB technique will still apply.

We close this section by noting that diffraction integral representations are also useful for incorporating symmetry constraints and boundary conditions. As an example of this, consider behavior near the saddle point of a double well model. We emphasized in Section 5.1 that if no bifurcation of the MPEP has occurred, the WKB tube of probability density centered on the axis will be well behaved as the saddle is approached. In particular, $w_2(x) = \partial^2 W/\partial y^2(x,0)$ will tend to $2|u_1(0)|$ as $x \to 0^+$. Since

$$\frac{\partial^2 W}{\partial x^2}(0,0) = \frac{\partial p_x}{\partial x}(0,0) = -2v_0'(0) \tag{6.11}$$

and $v_0(x) = u_x(x,0)$ is assumed to be smooth, in the absence of bifurcations W will to leading order be *locally quadratic* at the saddle. If $\boldsymbol{u}(x,y) \approx (\lambda_x x, -|\lambda_y|y)$ is the linearization of the drift at the saddle, we have $u_1(0) = -|\lambda_y|$ and $v_0'(0) = \lambda_x$. So, near $(x,y) = (0,0)$,

$$W(x,y) \approx W(0,0) - \lambda_x x^2 + |\lambda_y| y^2. \tag{6.12}$$

And

$$\widetilde{W}(p_x,p_y) \approx -W(0,0) - p_x^2/4\lambda_x + p_y^2/4|\lambda_y| \tag{6.13}$$

$$\widetilde{W}^{(x)}(p_x,y) \approx -W(0,0) - p_x^2/4\lambda_x - |\lambda_y| y^2 \tag{6.14}$$

will be the leading-order approximations to the Legendre-transformed actions.

Since the Hessian matrix $\partial^2\widetilde{W}/\partial p_i \partial p_j$ is not negative definite, an integral representation of the type (6.3) is not appropriate. But a representation of the type (6.5) may be used. In the absence of bifurcations K and $\widetilde{K}$ are well behaved near the saddle, so substituting (6.14) into (6.5) yields

$$\rho(x,y) \sim \text{const} \times \left\{ \epsilon^{-1/2} \int \exp\left[-(xp_x + p_x^2/4\lambda_x)/\epsilon\right] \, dp_x \right\} e^{-|\lambda_y|y^2/\epsilon}. \tag{6.15}$$

If p_x here is integrated from $-\infty$ to ∞, this approximation will be even in x. It will therefore serve as an approximation to the stationary density $\rho_0(x,y)$ near the saddle. The integral may be evaluated explicitly, and the approximation reduces to the standard inverted Gaussian approximation

$$\rho_0(x,y) \sim \text{const} \times e^{+\lambda_x x^2/\epsilon} e^{-|\lambda_y|y^2/\epsilon}. \tag{6.16}$$

But when approximating the *quasistationary* density $\rho_1(x,y)$ near the saddle, one needs an approximate solution of the Smoluchowski equation that is odd rather than even. Such an approximate solution is obtained by integrating p_x from 0 to ∞ rather than from $-\infty$ to ∞. If this is in fact done, eq. (6.15) reduces to (5.20), the standard Kramers-type error function approximation to the quasistationary density!

Although error function approximations originated (with Kramers) in an entirely different context, they fit naturally into the Maslov-WKB framework. We conclude that diffraction integral representations can be modified to incorporate the effects of symmetry constraints. In Section 8.3 we shall use a similar half-range integration in our integral representation for the quasistationary density near a nascent cusp.

7 Scaling Behavior Near a Cusp

We can apply the Maslov-WKB method of the last section to symmetric double well models in which the MPEP has bifurcated, and the instanton trajectories emerging from $S = (x_s, 0)$ focus at a point $(x_f, 0)$, with $0 < x_f < x_s$. As we shall see, behavior near the focal point $(x_f, 0)$ is best described in the language of critical phenomena.

The Maslov-WKB method was first applied to focusing (cusp) singularities in two-dimensional models by Dykman *et al.* [12]. Their analysis, which does not assume any sort of symmetry, specializes in the case of symmetry about the x-axis to the following. Assume that the Legendre-transformed action $\widetilde{W}^{(y)} = yp_y - W$, regarded as a function of x and p_y, may be asymptotically approximated near $(x, p_y) = (x_f, 0)$ by the cuspoid (codimension $n = 2$) normal form

$$\widetilde{W}^{(y)}(x, p_y) \sim -\frac{a_0}{4}p_y^4 - \frac{a_1}{2}(x - x_f)p_y^2 - w_0(x). \tag{7.1}$$

Here a_0 and a_1 are positive constants, and $w_0(x)$ is simply $W(x, 0)$, *i.e.*, $-\widetilde{W}^{(y)}(x, 0)$. Since $y(x, p_y) = \partial\widetilde{W}^{(y)}/\partial p_y(x, p_y)$, this assumption is equivalent to

$$y(x, p_y) \sim -a_0 p_y^3 - a_1(x - x_f)p_y \tag{7.2}$$

which is the phenomenological (Ginzburg-Landau) equation of state (5.33), discussed at length in Section 5.4. $\widetilde{W}^{(y)} = \widetilde{W}^{(y)}(x, p_y)$ can be viewed as a Helmholtz free energy, just as $W = W(x, y)$ can be viewed as a Gibbs free energy.

The cuspoid form for $\widetilde{W}^{(y)}$ is certainly consistent with the folding of the Lagrangian manifold $\mathcal{M}$, as seen (in projection) in Figs. 2(c) and 2(d). It is also consistent with the quantitative asymptotics of Section 5.4. Since $p_y = 0$ corresponds to $y = y(x, p_y) = 0$, (7.2) implies

$$\begin{aligned} w_2(x) = \frac{\partial^2 W}{\partial y^2}(x, 0) = \frac{\partial p_y}{\partial y}(x, 0) = \left[\frac{\partial y}{\partial p_y}(x, 0)\right]^{-1} \\ \sim -a_1^{-1}(x - x_f)^{-1}, \qquad x \to x_f^+. \end{aligned} \tag{7.3}$$

This is precisely the near-focus blowup behavior of eqs. (5.35)–(5.36), which we derived analytically from the Riccati equation (5.34). By comparing (7.3) with (5.34), we see that the constant a_1 must equal $1/v_0(x = x_f)$, the *reciprocal speed* of the on-axis instanton trajectory as it passes through the focus. Since $x - x_f$ is analogous to $T - T_c$ and w_2 to a (negative) magnetic susceptibility, the blowup of (7.3) is analogous to the critical exponent γ of the focus, in thermodynamic language, equalling unity.

Dykman *et al.* use a one-dimensional diffraction integral representation, resembling (6.5) but with x and y interchanged, to approximate the stationary probability density ρ_0 near $\boldsymbol{x} = (x_f, 0)$. A crucial assumption is that the transformed prefactor $\widetilde{K}^{(y)} = \widetilde{K}^{(y)}(x, p_y)$, which has no direct thermodynamic interpretation, is well behaved (locally constant, or 'slowly varying') near $(x, p_y) = (x_f, 0)$. If this is the case, and it may be approximated by a constant, one can construct the Maslov-WKB approximation

$$\begin{aligned} &K(x,y)\exp[-W(x,y)/\epsilon] \qquad (7.4)\\ &\quad\sim \epsilon^{-1/2}\int_{-\infty}^{\infty}\widetilde{K}^{(y)}(x,p_y)\exp\left\{\left[-yp_y+\widetilde{W}^{(y)}(x,p_y)\right]/\epsilon\right\}\,dp_y\\ &\quad\approx \epsilon^{-1/2}\widetilde{K}^{(y)}(x_f,0)e^{-W(x_f,0)/\epsilon}\int_{-\infty}^{\infty}\exp\left\{-\left[\frac{a_0}{4}p_y^4+\frac{a_1}{2}(x-x_f)p_y^2+yp_y\right]/\epsilon\right\}\,dp_y.\end{aligned}$$

In terms of 'stretched' variables $X \equiv (x - x_f)/\epsilon^{1/2}$ and $Y \equiv y/\epsilon^{3/4}$ this becomes

$$\epsilon^{-1/4}\widetilde{K}^{(y)}(x_f,0)e^{-W(x_f,0)/\epsilon}e^{-[XW'(x_f,0)/\epsilon^{1/2}+(X^2/2)W''(x_f,0)]}\,\mathfrak{P}(a_1a_0^{-1/2}X, a_0^{-1/4}Y), \qquad (7.5)$$

where the primes denote derivatives with respect to x. Here the canonical diffraction function

$$\mathfrak{P}(u,v) \equiv \int_{-\infty}^{\infty}\exp\left\{-\left[\frac{1}{4}t^4+\frac{1}{2}ut^2+vt\right]\right\}\,dt \qquad (7.6)$$

is a modified (real) Pearcey function (cf. Paris [35]).

The expression (7.5) is an asymptotically ($\epsilon \to 0$) valid approximation to the stationary density ρ_0 and quasistationary density ρ_1, on the $x - x_f = O(\epsilon^{1/2})$, $y = O(\epsilon^{3/4})$ lengthscale near the cusp $(x_f, 0)$. It supplements the WKB approximation, which is singular there. One sees that on this lengthscale, the pre-exponential factor in ρ_0 and ρ_1 is actually of magnitude $O(\epsilon^{-1/4})$. The *singularity index* of the cusp equals $1/4$, as in physical optics.

The absence of an 'i' from the exponent gives rise to unusual asymptotic behavior of the diffraction function. The familiar Pearcey fringes of physical optics are replaced by an exponential slope, which becomes increasingly steep as $\epsilon \to 0$. Beyond the cusp (*i.e.*, at $x < x_f$, which is analogous to $T < T_c$), the WKB approximation $K(x, y)\exp[-W(x, y)/\epsilon]$ is again valid, but W is no longer differentiable through the x-axis [28]. This is reflected in the far-field asymptotics of the Pearcey function $\mathfrak{P}$. One can show that in the far field, *i.e.*, as $X = (x - x_f)/\epsilon^{1/2} \to -\infty$, the expression (7.5) matches to a WKB approximation displaying this nondifferentiability. One can also show that the fold caustic emanating from $(x_f, 0)$, as in Fig. 2(c), is *nonphysical*. It arises from subdominant saddle points of the Pearcey integral, and does not contribute to the leading weak-noise asymptotics for ρ_0 and ρ_1. This is closely related to the fact that "optimal paths [*i.e.*, *physical* instanton trajectories] do not encounter caustics," as Dykman *et al.* [12] put it.

Now the preceding Maslov-WKB treatment is satisfactory so far as it goes. But it leaves unresolved the issue of the validity of the Ginzburg-Landau approximation. The quartic normal form (7.1) for $\widetilde{W}^{(y)} = \widetilde{W}^{(y)}(x, p_y)$, and the cubic equation of state (7.2) for its first derivative $y = y(x, p_y)$, model a second-order phase transition with *mean field* (*i.e.*, classical) *critical exponents*. Equivalently, they model the critical behavior of a system which, though it has a phase transition, has a *smooth thermodynamic surface*. In the present context, assuming the local validity of the Ginzburg-Landau approximation amounts to assuming that the Lagrangian manifold $\mathcal{M}$ is smooth through the point $(x, p_y) = (x_f, 0)$. Of course the surface turns vertical there, causing $\partial p_y/\partial y$ to diverge. The assumption is that the singularity can be transformed away by using x and p_y, rather than x and y, as independent variables.

This assumption requires proof. One could presumably justify it by analysing the smoothness (and blowup) properties of solutions of the Hamilton-Jacobi equation. But we shall give a different, more physical justification. First, we shall model the local behavior of W and $\widetilde{W}^{(y)}$ by a scaling law, as in the modern theory of critical phenomena. Our treatment will serve as a warmup for Section 8, where we shall analyse the much more complicated (nonclassical) singularity appearing in models where the MPEP is beginning to bifurcate.

To see that a scaling law is appropriate in models with a bifurcated MPEP, consider the behavior of the on-axis transverse derivatives $w_{2m} = \partial^{2m}W/\partial x^{2m}(x, 0)$ as $x \to x_f^+$. We know by (5.35)–(5.36) that w_2 diverges as $(x - x_f)^{-1}$. The Riccati equation satisfied by w_2 is only the first of a hierarchy of ordinary differential equations, describing the evolution of the functions w_{2m} as one moves along the on-axis instanton trajectory from S (where $x = x_s$, and $t = -\infty$) to the saddle (where $x = 0$, and $t = +\infty$). For example, w_4 satisfies the ODE (5.15). w_2 appears in each of the higher equations, and its blowup will induce a blowup of $w_4, w_6, \ldots$ It is not difficult to show that

$$w_{2m}(x) = \frac{\partial^{2m}W}{\partial x^{2m}}(x, 0) \sim \text{const} \times (x - x_f)^{-(3m-2)}, \qquad x \to x_f^+, \tag{7.7}$$

for $2m = 2, 4, 6, \ldots$ These blowup rates motivate the *scaling Ansatz*

$$W(x, y) \sim W(x, 0) + (x - x_f)^2 h_\pm\left(\frac{y}{|x - x_f|^{3/2}}\right), \qquad x \to x_f^\pm \tag{7.8}$$

for the behavior of W near the cusp $(x_f, 0)$. Here the exponents 2 and $3/2$ are determined uniquely by the m-dependence of the blowup rates, and the functions $h_\pm(\cdot)$ of the scaling variable $z \equiv y/|x - x_f|^{3/2}$ are not yet determined (though they must be even). This Ansatz is assumed to be accurate to $O((x - x_f)^2)$, when $y = O(|x - x_f|^{3/2})$. We could equally well posit

$$W\left(x, z|x - x_f|^{3/2}\right) \sim W(x_f, 0) + (x - x_f)W'(x_f, 0) + \frac{(x - x_f)^2}{2}W''(x_f, 0) + (x - x_f)^2 h_\pm(z) \tag{7.9}$$

as $x \to x_f^\pm$, since we are assuming the accuracy of the scaling Ansatz only up to $O((x - x_f)^2)$. The first three terms in this asymptotic approximation are 'regular'; the scaling behavior appears only in the final, singular term.

The exponents 2 and $3/2$ are typical of a mean field theory. One can show that the scaling functions $h_\pm$ are also those of a mean field theory. They may be computed by substituting the scaling Ansatz (7.9) into the Hamilton-Jacobi equation $H(\boldsymbol{x}, \boldsymbol{\nabla} W) = 0$. For this, one needs to rewrite the Hamilton-Jacobi equation in terms of the independent variables x and z. Using the formula (5.2) for H, and the expansions (5.9), one finds

$$\begin{aligned} H(\boldsymbol{x},\boldsymbol{p}) &= \frac{p_x^2}{2} + \frac{p_y^2}{2} + u_x(x,y)p_x + u_y(x,y)p_y \\ &= \frac{p_x^2}{2} + \frac{p_y^2}{2} + \Big[v_0(x) + v_2(x)y^2 + \cdots\Big]\,p_x + \Big[u_1(x)y + u_3y^3 + \cdots\Big]\,p_y \\ &\approx \frac{p_x^2}{2} + \frac{p_y^2}{2} + \Big[v_0(x_f) \pm v_0'(x_f)\,|x-x_f|\Big]\,p_x \end{aligned} \tag{7.10}$$

up to $O(|x-x_f|^1)$ accuracy, since $y = z|x-x_f|^{3/2}$. It follows from the scaling form (7.8) that up to $O(|x-x_f|^1)$ accuracy,

$$\begin{aligned} p_x(x,y) = \frac{\partial W}{\partial x}(x,y) &\sim -2v_0(x) \pm \Big[2h_\pm(z) - (3/2)zh_\pm'(z)\Big]\,|x-x_f| \\ &\approx -2\Big[v_0(x_f) \pm v_0'(x_f)\,|x-x_f|\Big] \pm \Big[2h_\pm(z) - (3/2)zh_\pm'(z)\Big]\,|x-x_f| \end{aligned} \tag{7.11}$$

$$p_y(x,y) = \frac{\partial W}{\partial y}(x,y) \sim |x-x_f|^{1/2}\,h_\pm'(z) \tag{7.12}$$

where we have used the fact (see Section 5.2) that $W'(x,0) = w_0'(x) = -2v_0(x)$. Substituting (7.11)–(7.12) into (7.10), and setting the coefficient of $|x-x_f|^1$ equal to zero, yields the ODE

$$(h_\pm')^2 = \pm v_0(x_f)\,\Big[4h_\pm - 3zh_\pm'\Big]. \tag{7.13}$$

It is easier to solve for z as a function of $h_\pm'$, than for $h_\pm$ as a function of z. One finds

$$z = z(h_\pm') = -C(h_\pm')^3 \mp v_0(x_f)^{-1}h_\pm' \tag{7.14}$$

where C is undetermined. But $z = y/|x-x_f|^{3/2}$, and by (7.12), $h_\pm' \sim p_y/|x-x_f|^{1/2}$. Rewriting z in terms of y and $|x-x_f|$, and h' in terms of p_y and $|x-x_f|$, yields

$$\begin{aligned} y = y(x,p_y) &\sim -Cp_y^3 \mp v_0(x_f)^{-1}\,|x-x_f|\,p_y \\ &= -Cp_y^3 - v_0(x_f)^{-1}(x-x_f)p_y. \end{aligned} \tag{7.15}$$

If one identifies the model-dependent constant C with a_0, this is precisely eq. (7.2), the mean field (Ginzburg-Landau) equation of state! It is valid on both sides of the on-axis focus, *i.e.*, both when $x - x_f > 0$ and when $x - x_f < 0$.

This derivation illustrates how one may go from the pattern of blowup rates of the transverse derivatives $w_{2m}(x) = \partial^{2m}W/\partial y^{2m}(x,0)$ as $(x_f,0)$ is approached, to a scaling form for W, to an equation of state. The singular behavior of the WKB prefactor K can be analysed similarly (we only summarize the analysis). We know by (5.36) that $K(x,0)$ diverges as $(x-x_f)^{-1/2}$ when $x \to x_f^+$. A scaling form

$$K(x,y) \sim \text{const} \times |x-x_f|^{-1/2}\,q_\pm\left(\frac{y}{|x-x_f|^{3/2}}\right), \qquad x \to x_f^\pm, \tag{7.16}$$

modelled after the scaling form (7.8) for W, may be used to approximate K away from the x-axis. This approximation should be accurate to $O(|x - x_f|^{-1/2})$ as $x \to x_f^{\pm}$, when $y = O(|x - x_f|^{3/2})$. By substituting the two scaling forms (7.8) and (7.16) into the transport equation (5.7) for K, and working to leading order near $(x_f, 0)$, one can determine the scaling functions $q_{\pm} = q_{\pm}(z)$. It is easily verified that collecting the $O(|x - x_f|^{-3/2})$ terms in the transport equation yields the ODE

$$\left[2h'_{\pm} \pm 3v_0(x_f)z\right] q'_{\pm} + \left[h''_{\pm} \pm v_0(x_f)\right] q_{\pm} = 0, \tag{7.17}$$

which $q_{\pm} = q_{\pm}(z)$ must satisfy. Using elementary calculus, and the fact that $h_{\pm} = h_{\pm}(z)$ satisfies the ODE (7.13), one can show that eq. (7.17) has solution

$$q_{\pm}(z) = \text{const} \times \sqrt{-h''_{\pm}(z)}. \tag{7.18}$$

But since $z = y/|x - x_f|^{3/2}$, we know by (7.12) that

$$h''_{\pm}(z) \sim |x - x_f| \frac{\partial p_y}{\partial y} \left(x, y = z|x - x_f|^{3/2}\right). \tag{7.19}$$

Substituting (7.18) and (7.19) into the scaling form (7.16) for K reduces it to

$$K(x, y) \sim \text{const} \times \sqrt{-\frac{\partial p_y}{\partial y}(x, y)}. \tag{7.20}$$

This asymptotic approximation is very simple, and has a profound consequence. We know that the transformed prefactor $\widetilde{K}^{(y)}(x, p_y)$ can be obtained from $K(x, y)$ by dividing by a 'Van Vleck factor,' as in (6.6). We therefore have that

$$\widetilde{K}^{(y)}(x, p_y) \propto K(x, y) \Bigg/ \sqrt{-\frac{\partial^2 W}{\partial y^2}(x, y)} \tag{7.21}$$

$$\sim \text{const}, \tag{7.22}$$

since $\partial^2 W/\partial y^2 = \partial p_y/\partial y$. This *constant* asymptotic approximation is accurate to leading order as $x \to x_f$, when $y = O(|x - x_f|^{3/2})$.

We have just deduced that on the appropriate lengthscale near the focus, *i.e.*, $x - x_f = o(1)$ and $y = O(|x - x_f|^{3/2})$, *the transformed WKB prefactor* $\widetilde{K}^{(y)}$ *does not diverge*. $\widetilde{K}^{(y)}$, unlike the prefactor K itself, is asymptotically constant near the focus. This was the crucial assumption made by Dykman *et al.*, and we see that like the Ginzburg-Landau normal form for the Legendre-transformed action, it is justified by our scaling theory of local behavior.

We conclude that at least in the case of a generic (cusp) singularity, by investigating the blowup rates of the transverse action derivatives as the singularity is approached, one can derive scaling relations for W and K, and ultimately construct a Maslov-WKB approximation to the stationary probability density near the singularity. This technique is not restricted to singularities of the classical Ginzburg-Landau type.

8 Scaling Behavior Near a Nascent Cusp

Finally, we can construct a scaling theory of weak-noise behavior near the 'nascent cusp' singularity appearing at the saddle point of any symmetric double well model, at the onset of bifurcation. The construction will closely parallel the construction of the last section. But several novel features will appear. We shall find that Legendre-transformed versions of the action are approximated, in the vicinity of a nascent cusp, by *nonpolynomial* normal forms. Equivalently, the nascent cusp singularity, unlike an on-axis focus, will prove to have *nonclassical* critical exponents. The exponents will depend continuously on the parameter $\mu \equiv |\lambda_y|/\lambda_x$, which characterizes the linearized drift field at the saddle.

The universal presence at criticality of a nongeneric two-sided caustic (which, as shown in Fig. 4, extends sideways from the saddle point) will follow from the nonpolynomial normal forms for the Legendre-transformed actions. Indeed, one of the normal forms will supply a *nonpolynomial unfolding* of the nongeneric caustic. Moreover, the fact that the critical exponents of the nascent cusp are model-dependent and continuously varying will induce a continuously varying singularity index, and a continuously varying prefactor exponent in the non-Arrhenius weak-noise MFPT asymptotics. To see this, we shall have to go beyond the WKB approximation, by applying the Maslov-WKB method. In Section 8.1 we analyse the scaling properties of the action and the WKB prefactor, and in Section 8.2, we compare our scaling formulæ with numerical data. In Section 8.3 we apply the Maslov-WKB method, and construct weak-noise approximations to the stationary and quasistationary probability densities near the saddle.

8.1 Scaling in the WKB Approximation

Our scaling treatment of the nascent cusp begins with an investigation of the blowup rates of the transverse action derivatives $w_{2m}(x) = \partial^{2m}W/\partial y^{2m}(x,0)$, as $x \to 0^+$. Up to now we have written down only the ODE's satisfied by w_2 (*i.e.*, the Riccati equation (5.14)) and w_4 (*i.e.*, eq. (5.15)). The full hierarchy of ODE's may be derived by substituting the Taylor series $\sum_{m=0}^{\infty} w_{2m}(x)y^{2m}/(2m)!$ for $W(x,y)$ into the Hamilton-Jacobi equation $H(\boldsymbol{x}, \boldsymbol{\nabla}W) = 0$, and separating out the coefficients of each power of y. One finds

$$\begin{aligned} \dot{w}_{2m} &= -v_0 w_{2m}' & (8.1)\\ &= -\sum_{k=1}^{m} \binom{2m}{2k-1} [w_{2k}/2 + \tilde{u}_{2k-1}] w_{2m-2k+2} - \sum_{j=1}^{m-1} \binom{2m}{2j} [w_{2j}'/2 + \tilde{v}_{2j}] w_{2m-2j}' + 2\tilde{v}_0\tilde{v}_{2m}, \end{aligned}$$

where $\tilde{u}_{2j+1} \equiv (2j+1)!\, u_{2j+1}$ and $\tilde{v}_{2j} \equiv (2j)!\, v_{2j}$, and u_{2j+1} and v_{2j} are the drift velocity derivatives defined in (5.9). As usual, the time derivative here is with respect to transit time of the on-axis instanton trajectory, which satisfies $\dot{x} = -v_0(x)$ as it moves from $S = (x_s, 0)$ to the saddle. Since $v_0(x) \equiv u_x(x,0)$, this trajectory $t \mapsto x^*(t)$ moves anti-parallel to the drift. And since $\boldsymbol{u}(x,y) \approx (\lambda_x x, -|\lambda_y|y)$ near $(0,0)$, $x^*(t)$ is approximated (as $t \to +\infty$) by const $\times\, e^{-\lambda_x t}$.

We showed in Section 5.4, by analysing the Jacobi equation satisfied by the transverse soft mode, that

$$w_2(x) \sim \text{const} \times x^2, \quad x \to 0^+ \qquad (8.2)$$

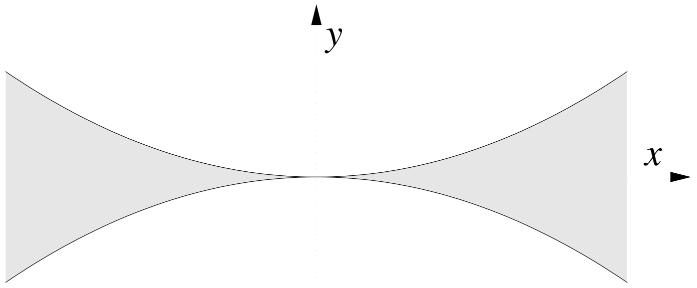

Figure 8: A sketch of the near-axis Region N, defined by $|y|/|x|^{2\mu} \le$ const. The choice of constant is immaterial, so long as it is positive. Most of our expansions for the action W and its Legendre transforms, in double well models at criticality, are valid as $(x, y) \to (0,0)$ from within Region N. The region could equally well be defined by $|p_y|/|x|^{2\mu} \le$ const, $|y|/|p_x|^{2\mu} \le$ const, or $|p_y|/|p_x|^{2\mu} \le$ const. To leading order near $(0,0)$, the four definitions are equivalent.

in any double well model at the onset of bifurcation (see eqs. (5.54)–(5.55), and Fig. 7). If the fact that $w_2 \to 0$ as $x \to 0^+$ (*i.e.*, as $t \to +\infty$) is substituted into the general ODE (8.1), it is easy to show, by integrating forward in time toward $t = +\infty$, that

$$w_{2m}(x) \sim \text{const} \times x^{-(4m-4)\mu}, \quad x \to 0^+, \tag{8.3}$$

for $2m = 4, 6, 8, \ldots$ This pattern of blowup rates, as the saddle is approached, motivates the scaling Ansatz (cf. (7.8))

$$W(x, y) \sim W(x, 0) + |x|^{4\mu} h(y/|x|^{2\mu}), \qquad x \to 0. \tag{8.4}$$

Here $h(\cdot)$ is some (even) scaling function, as yet undetermined, and the exponents 4μ and 2μ are determined uniquely by the m-dependence of the blowup rates of (8.3). This Ansatz is assumed to be accurate to $O(|x|^{4\mu})$ as $x \to 0$, when $y = O(|x|^{2\mu})$. We could equally well posit a finite-length asymptotic expansion for $W(x, z|x|^{2\mu})$, namely

$$W(x, z|x|^{2\mu}) \sim \left(\sum_{k=0}^{\lfloor 4\mu \rfloor} \frac{\partial^k W}{\partial x^k}(0,0) \frac{x^k}{k!} \right) + |x|^{4\mu} h(z), \qquad x \to 0. \tag{8.5}$$

Here $z \equiv y/|x|^{2\mu}$ is the scaling variable. Only even powers of x appear in the summation, and by convention, here and below $\lfloor 4\mu \rfloor$ denotes the greatest *even* integer less than or equal to 4μ. The expansion (8.5) is assumed to be accurate to $O(|x|^{4\mu})$, at any fixed value of z. It can be thought of as an asymptotic development of $W(\boldsymbol{x})$, as $\boldsymbol{x} \to \mathbf{0}$ from within the near-axis region defined by the condition $|z| \le$ const. This condition defines a notch-shaped region, which we call Region N. (See Fig. 8.)

It follows by differentiating (8.4) twice with respect to y that $w_2(x) \sim h''(0)$ as $x \to 0$. For consistency with the 'splayout' behavior $w_2(x) \sim \text{const} \times x^2$ of (8.2), we must have $h''(0) = 0$. Notice the slight discrepancy: the falloff rate of w_2 is not fully captured by the scaling Ansatz. Actually, this is unsurprising. A term proportional to x^2y^2 in W, such as would arise from the $O(x^2)$ falloff of w_2 as $x \to 0$, would (in terms of x and $z = y/|x|^{2\mu}$) be proportional to $|x|^{4\mu+2}z^2$. It would therefore be negligible in comparison to the scaling term $|x|^{4\mu}h(z)$, as $x \to 0$. The scaling term captures the blowup as $x \to 0$ of w_4, w_6, $w_8, \ldots$, but capturing the precise falloff rate of w_2 would require a more refined analysis. We shall not attempt to include in our Ansatz the 'sub-scaling' terms that such an analysis would require.

The scaling function $h(\cdot)$ may be computed by the technique used in Section 7. By substituting the expression (8.5) into the Hamilton-Jacobi equation $H(\boldsymbol{x}, \boldsymbol{\nabla} W) = 0$, rewriting the Hamilton-Jacobi equation in terms of the independent variables x and z, and setting the coefficient of $|x|^{4\mu}$ equal to zero, one obtains an ODE for $h = h(z)$. This ODE turns out to be (cf. (7.13))

$$(h')^2 = 2\,|\lambda_y|\,[4h - zh']\,. \tag{8.6}$$

As in Section 7, it is easier to solve for z as a function of h', than for h as a function of z. One finds (cf. (7.14))

$$z = z(h') = h'/2|\lambda_y| + c(h')^{1/3}, \tag{8.7}$$

where c is undetermined. But $z = y/|x|^{2\mu}$ and $p_y = \partial W/\partial y \sim |x|^{2\mu}h'(z)$. Rewriting z in terms of x and y, and h' in terms of x and p_y, yields an asymptotically accurate equation of state (cf. (7.15))

$$y = y(x, p_y) \sim p_y/2|\lambda_y| + c_{y;x,p_y}|x|^{4\mu/3}p_y^{1/3}, \tag{8.8}$$

where $c_{y;x,p_y} = c$. In practice the model-dependent constant c would be computed numerically, by fitting (8.8) to the flow field of instanton trajectories in the vicinity of the saddle. We must have $c > 0$, since the map $p_y \mapsto y$ is necessarily monotone increasing near the saddle. This is because 'whorling,' as in Fig. 6, occurs only in models with a bifurcated MPEP. Whorling is absent at criticality, *i.e.*, at the onset of bifurcation.

The equation of state (8.8) is certainly not of the classical Ginzburg-Landau form. By anti-differentiating it, we can obtain an equally unusual approximation to the Legendre-transformed action $\widetilde{W}^{(y)} = yp_y - W$, where $p_y = \partial W/\partial y$. Since $y = \partial\widetilde{W}^{(y)}/\partial p_y$, we necessarily have

$$\widetilde{W}^{(y)}(x, p_y) \sim -W(x, 0) + p_y^2/4|\lambda_y| + C_{x,p_y}|x|^{4\mu/3}p_y^{4/3} \tag{8.9}$$

$$\sim -\left(\sum_{k=0}^{\lfloor 4\mu \rfloor} \frac{\partial^k W}{\partial x^k}(0,0)\frac{x^k}{k!}\right) + p_y^2/4|\lambda_y| + C_{x,p_y}|x|^{4\mu/3}p_y^{4/3}, \tag{8.10}$$

where $C_{x,p_y} = 3c/4$. This asymptotic approximation should be accurate to $O(|x|^{4\mu})$, when $p_y = O(|x|^{2\mu})$ [*i.e.*, when $y = O(|x|^{2\mu})$, or when $\boldsymbol{x} \to \boldsymbol{0}$ from within Region N].

The formula (8.10) can be called a *nonpolynomial normal form* for the transformed action $\widetilde{W}^{(y)}$ near the nascent cusp. Notice that as $\boldsymbol{x} \to \boldsymbol{0}$, the final, nonpolynomial term $C_{x,p_y}|x|^{4\mu/3}p_y^{4/3}$ is significant in a relative sense only within Region N. In the far field of the $p_y = O(|x|^{2\mu})$ lengthscale, as $x \to 0$ it is increasingly dominated by the p_y^2 term, and the normal form reduces to a polynomial. The $C_{x,p_y}|x|^{4\mu/3}p_y^{4/3}$ term plays a much more important role in the near field. One can think of (8.10) as providing an *interpolation* between the non-polynomial asymptotic development that is valid as $\boldsymbol{x} \to \mathbf{0}$ from within Region N, and the polynomial development that is valid as $\boldsymbol{x} \to \mathbf{0}$ from within its far field. *The scaling behavior is visible only within Region N.*

It is worth noting that despite its asymptotic validity, the nonpolynomial normal form (8.10) does not fully capture the $p_y \to 0$ behavior of $\widetilde{W}^{(y)}(x, p_y)$ at *fixed*, nonzero x. If the nonanalytic $p_y^{4/3}$ falloff were exact, it would follow by differentiating twice with respect to p_y that $\partial^2\widetilde{W}^{(y)}/\partial p_y^2$, *i.e.*, $\partial y/\partial p_y$, would diverge as $p_y \to 0$. This would imply that $w_2 = \partial p_y/\partial y(y = 0)$ would be *identically zero* at any nonzero x. But we know that $w_2(x) \sim \text{const} \times x^2$, $x \to 0$. The discrepancy is due to the fact that the nonzero w_2 near $x = 0$ arises from 'sub-scaling' behavior that we are not attempting to model. It is not difficult to see that at fixed nonzero x, the apparent nonanalyticity at $p_y = 0$ must be 'rounded' at a lengthscale $p_y = O(|x|^{2\mu+3})$, or equivalently at $y = O(|x|^{2\mu+1})$, to yield consistency with the $w_2(x) \sim \text{const} \times x^2$ asymptotics. However, on the $p_y = O(|x|^{2\mu})$ lengthscale the rounding becomes invisible as $x \to 0$.

With its continuously varying (in general, irrational) exponent $4\mu/3$, the normal form for $\widetilde{W}^{(y)}$ in Region N looks quite different from the normal forms of catastrophe theory [2, 4]. Its most striking feature is the non-analyticity at $(x, p_y) = (0, 0)$, which can be interpreted thermodynamically. Recall that $\widetilde{W}^{(y)}$ (like $\widetilde{W}^{(x)}$, $\widetilde{W}$, and W) can be viewed as a *thermodynamic potential* on the thermodynamic surface (*i.e.*, Lagrangian manifold) $\mathcal{M}$. In fact, through its derivatives it determines the shape of $\mathcal{M}$. So at $(x, p_y) = (0, 0)$, or equivalently at $(x, y) = (0, 0)$, the surface $\mathcal{M}$ will itself be non-analytic. However, as μ increases, $\widetilde{W}^{(y)}$ becomes increasingly differentiable (with respect to x, at least) at $x = 0$. The order of the 'phase transition' appearing at the saddle at criticality is, therefore, an increasing function of μ.

We can Legendre-transform the normal form for $\widetilde{W}^{(y)}$ to obtain a normal form for the double Legendre transform $\widetilde{W} = \boldsymbol{x} \cdot \boldsymbol{p} - W = xp_x + \widetilde{W}^{(y)}$, as a function of p_x and p_y. A further Legendre transform will yield a normal form for the remaining thermodynamic potential, $\widetilde{W}^{(x)}$. We sketch only the first of these two computations. Differentiating (8.9)–(8.10) with respect to x, and using $p_x = \partial W/\partial x = -\partial\widetilde{W}^{(y)}/\partial x$, yields

$$p_x = p_x(x, p_y) \sim p_x(x, 0) + c_{p_x;x,p_y}\left[|x|^{4\mu/3-1}\,\text{sgn}\, x\right] p_y^{4/3} \tag{8.11}$$

$$\sim \left(-2\lambda_x x + \cdots + \text{const} \times x^{\lfloor 4\mu \rfloor - 1}\right) + c_{p_x;x,p_y}\left[|x|^{4\mu/3-1}\,\text{sgn}\, x\right] p_y^{4/3}, \tag{8.12}$$

where $c_{p_x;x,p_y} = -(4\mu/3)C_{x,p_y} = -\mu c$. Here we have used the fact that $W(x, 0) = -2v_0(x)$, so that $W''(0, 0) = -2\lambda_x$, etc. This approximation is accurate to $O(|x|^{4\mu-1})$ as $x \to 0$, when $p_y = O(|x|^{2\mu})$ [*i.e.*, when $\boldsymbol{x} \to \mathbf{0}$ from within Region N]. If $\mu > 1/2$, it is easy to invert the series (8.12) to approximate $x = x(p_x, p_y)$. The $-2\lambda_x x$ term is dominant, and inversion yields

$$x = x(p_x, p_y) \sim x(p_x, 0) + c_{x;p_x,p_y} \left[|p_x|^{4\mu/3-1} \operatorname{sgn} p_x \right] p_y^{4/3} \tag{8.13}$$

$$\sim \left(-p_x/2\lambda_x + \cdots + \text{const} \times p_x^{\lfloor 4\mu \rfloor - 1} \right) + c_{x;p_x,p_y} \left[|p_x|^{4\mu/3-1} \operatorname{sgn} p_x \right] p_y^{4/3}, \tag{8.14}$$

where $c_{x;p_x,p_y} = -(2\lambda_x)^{-4\mu/3} c_{p_x;x,p_y} = (2\lambda_x)^{-4\mu/3}\mu c$. Since $x(p_x, p_y) = \partial\widetilde{W}/\partial p_x(p_x, p_y)$, we must have

$$\widetilde{W}(p_x, p_y) \sim \widetilde{W}(p_x, 0) + p_y^2/4|\lambda_y| + C_{p_x,p_y}|p_x|^{4\mu/3} p_y^{4/3} \tag{8.15}$$

$$\sim \left(-W(0,0) - p_x^2/4\lambda_x + \cdots + \text{const} \times p_x^{\lfloor 4\mu \rfloor} \right) + p_y^2/4|\lambda_y| + C_{p_x,p_y}|p_x|^{4\mu/3} p_y^{4/3} \tag{8.16}$$

$$\approx -W(0,0) - p_x^2/4\lambda_x + p_y^2/4|\lambda_y| + C_{p_x,p_y}|p_x|^{4\mu/3} p_y^{4/3}, \tag{8.17}$$

where $C_{p_x,p_y} = (3/4\mu) c_{x;p_x,p_y} = (3c/4)(2\lambda_x)^{-4\mu/3}$. The momentum-space normal forms (8.15) and (8.16) should be accurate to $O(|p_x|^{4\mu})$ as $p_x \to 0$, when $p_y = O(|p_x|^{2\mu})$. This is simply a momentum-space version of the condition that $\boldsymbol{x} \to \boldsymbol{0}$ from within Region N.

It is useful to compare the truncated normal form (8.17) with (6.13), the quadratic approximation to $\widetilde{W}$ that is valid near the saddle point in the absence of focusing. We see that the fact that a double well model is 'critical' modifies the double Legendre transform $\widetilde{W}$ near the saddle in a very simple way: it adds the final, nonpolynomial term. In a sense, the coefficient C_{p_x,p_y} measures the *strength* of the nascent cusp singularity at the saddle.

The computation of the remaining thermodynamic potential, $\widetilde{W}^{(x)}$, is left to the reader. In Figure 9 we list the normal forms for $\widetilde{W}$, $\widetilde{W}^{(x)}$, and $\widetilde{W}^{(y)}$, as well as the scaling form for W. The expressions listed there are accurate to $O(|x|^{2\mu})$, *i.e.*, to $O(|p_x|^{2\mu})$, as the nascent cusp is approached from within Region N. In Figure 10 we list the four possible equations of state for x and y. They are accurate to $O(|x|^{2\mu-1})$, *i.e.*, to $O(|p_x|^{2\mu-1})$, in the same limit.

We emphasize that the normal form (8.16) for $\widetilde{W}$, the normal form for $\widetilde{W}^{(x)}$, and the equations of state that follow from them, are valid only for critical models with $\mu > 1/2$. The reason is that when $\mu \leq 1/2$, the final term in (8.12), which when $p_y = O(|x|^{2\mu})$ is of magnitude $O(x^{4\mu-1})$, is at least as large as the $-2\lambda_x x$ term as $x \to 0$. In fact when $\mu < 1/2$, in Region N (except on the x-axis) the leading asymptotics of $p_x = p_x(x, p_y)$ are not linear in x. This makes difficult the computation of asymptotic approximations to $x = x(p_x, p_y)$ and $\widetilde{W} = \widetilde{W}(p_x, p_y)$. For this reason we shall assume $\mu > 1/2$ henceforth.

It is a reasonable conjecture that in critical models where the symmetrical approximation (8.17) to $\widetilde{W} = \widetilde{W}(p_x, p_y)$ is valid, it is valid not merely near the p_x-axis (*i.e.*, in Region N), but uniformly as $\boldsymbol{p} \to \boldsymbol{0}$. One would like to substitute it into the Maslov-WKB diffraction integral (6.3), so as to obtain boundary layer approximations to the stationary and quasistationary probability densities near the saddle point $(0,0)$. The approximation to the quasistationary density would be a replacement for the usual Kramers-type error function approximation, (5.21). From it, one could derive an Eyring formula for the MFPT asymptotics, as in Section 5.2. Unfortunately there is a problem. If $C_{p_x,p_y} = 0$ and (8.17) becomes quadratic, the Hessian matrix $\partial^2\widetilde{W}/\partial p_i \partial p_j$ is clearly not

$$
\begin{aligned}
\widetilde{W}(p_x,p_y) &\sim \widetilde{W}(p_x,0) + p_y^2/4|\lambda_y| + C_{p_x,p_y}|p_x|^{4\mu/3}p_y^{4/3} \\
&\sim \Big(-W(0,0) - p_x^2/4\lambda_x + \cdots + \text{const}\times p_x^{\lfloor 4\mu\rfloor}\Big) + p_y^2/4|\lambda_y| + C_{p_x,p_y}|p_x|^{4\mu/3}p_y^{4/3} \\
\widetilde{W}^{(x)}(p_x,y) &\sim \widetilde{W}^{(x)}(p_x,0) - |\lambda_y|y^2 + C_{p_x,y}|p_x|^{4\mu/3}y^{4/3} \\
&\sim \Big(-W(0,0) - p_x^2/4\lambda_x + \cdots + \text{const}\times p_x^{\lfloor 4\mu\rfloor}\Big) - |\lambda_y|y^2 + C_{p_x,y}|p_x|^{4\mu/3}y^{4/3} \\
\widetilde{W}^{(y)}(x,p_y) &\sim \widetilde{W}^{(y)}(x,0) + p_y^2/4|\lambda_y| + C_{x,p_y}|x|^{4\mu/3}p_y^{4/3} \\
&\sim \Big(-W(0,0) + \lambda_x x^2 + \cdots + \text{const}\times x^{\lfloor 4\mu\rfloor}\Big) + p_y^2/4|\lambda_y| + C_{x,p_y}|x|^{4\mu/3}p_y^{4/3} \\
W(x,y) &\sim W(x,0) + |x|^{4\mu}h(y/|x|^{2\mu}) \\
&\sim \Big(W(0,0) - \lambda_x x^2 + \cdots + \text{const}\times x^{\lfloor 4\mu\rfloor}\Big) + |x|^{4\mu}h(y/|x|^{2\mu})
\end{aligned}
$$

Figure 9: Normal forms for the thermodynamic potentials (the Legendre transforms of the action W, and W itself) in the vicinity of a nascent cusp. In terms of the model-dependent constant c, $C_{p_x,p_y} = (3c/4)(2\lambda_x)^{-4\mu/3}$, $C_{p_x,y} = (3c/4)(2\lambda_x)^{-4\mu/3}(2|\lambda_y|)^{4/3}$, and $C_{x,p_y} = 3c/4$. These asymptotic expansions are valid in Region N, in critical double well models with $\mu > 1/2$. They are accurate to $O(|x|^{4\mu})$, or equivalently to $O(|p_x|^{4\mu})$.

$$
\begin{aligned}
x(p_x,y) &\sim x(p_x,0) + c_{x;p_x,y}\left[|p_x|^{4\mu/3-1}\,\text{sgn}\,p_x\right]y^{4/3} \\
&\sim \Big(-p_x/2\lambda_x + \cdots + \text{const}\times p_x^{\lfloor 4\mu\rfloor-1}\Big) + c_{x;p_x,y}\left[|p_x|^{4\mu/3-1}\,\text{sgn}\,p_x\right]y^{4/3} \\
x(p_x,p_y) &\sim x(p_x,0) + c_{x;p_x,p_y}\left[|p_x|^{4\mu/3-1}\,\text{sgn}\,p_x\right]p_y^{4/3} \\
&\sim \Big(-p_x/2\lambda_x + \cdots + \text{const}\times p_x^{\lfloor 4\mu\rfloor-1}\Big) + c_{x;p_x,p_y}\left[|p_x|^{4\mu/3-1}\,\text{sgn}\,p_x\right]p_y^{4/3} \\
y(x,p_y) &\sim p_y/2\,|\lambda_y| + c_{y;x,p_y}\,|x|^{4\mu/3}\,p_y^{1/3} \\
y(p_x,p_y) &\sim p_y/2\,|\lambda_y| + c_{y;p_x,p_y}\,|p_x|^{4\mu/3}\,p_y^{1/3}
\end{aligned}
$$

Figure 10: The four asymptotic equations of state, which describe the shape of the Lagrangian manifold $\mathcal{M}$ in the vicinity of a nascent cusp. It follows by differentiating the normal forms listed in Fig. 9 that $c_{x;I,J} = (4\mu/3)C_{I,J}$ and $c_{y;I,J} = (4/3)C_{I,J}$. So $c_{x;p_x,p_y} = \mu c/(2\lambda_x)^{4\mu/3}$, $c_{x;p_x,y} = \mu c(2|\lambda_y|)^{4/3}/(2\lambda_x)^{4\mu/3}$, $c_{y;p_x,p_y} = c/(2\lambda_x)^{4\mu/3}$, and $c_{y;x,p_y} = c$. These asymptotic expansions are valid in Region N, in critical double well models with $\mu > 1/2$.

negative definite. As we noted in Section 6, this precludes the use of the two-dimensional diffraction integral (6.3). The situation does not improve much if C_{p_x,p_y} is positive, so it is preferable to use an alternative integral representation. The asymptotic approximation to the Legendre transform $\widetilde{W}^{(x)} = xp_x - W = -yp_y + \widetilde{W}$, as a function of p_x and y, is listed in the table in Figure 9. A truncated version of it would be

$$\widetilde{W}^{(x)}(p_x, y) \approx -W(0,0) - p_x^2/4\lambda_x - |\lambda_y|\, y^2 + C_{y,p_x}|p_x|^{4\mu/3}y^{4/3}, \tag{8.18}$$

which is a nonpolynomial modification of the Gaussian approximation (6.14). This approximation is precisely what is needed in the one-dimensional diffraction integral (6.5), which is what we shall use instead of (6.3).

The reader may wonder about the domain of validity of the approximation (8.18) to $\widetilde{W}^{(x)} = \widetilde{W}^{(x)}(p_x, y)$. Is it valid outside Region N? In Section 8.2 we present numerical evidence that it is, in fact, a useful asymptotic approximation near the y-axis, even at fixed, nonzero y. Indeed, it explains the mysterious 'sideways' caustic of Fig. 4! To see this, differentiate (8.18) with respect to p_x to get

$$x = x(p_x, y) \approx -p_x/2\lambda_x + c_{x;p_x,y}\, y^{4/3}\Big[|p_x|^{(4\mu/3)-1}\,\mathrm{sgn}\, p_x\Big], \tag{8.19}$$

which is a truncated version of the asymptotic expansion of $x = x(p_x, y)$ listed in Figure 10. If $\mu < 3/2$, the formula (8.19) predicts that at any nonzero y, the map $p_x \mapsto x$ will *not be monotone*. This is because the coefficient $c_{x;p_x,y}$ is positive. By examination, if

$$|x| \lesssim \text{const} \times |y|^{(3/2-\mu)^{-1}}, \quad y \to 0, \tag{8.20}$$

then the inverse map $p_x = p_x(x, y)$ (and hence $W = W(x, y)$) will be multivalued. This inequality defines a two-sided *nongeneric caustic*, which emanates from $(0, 0)$ along the positive and negative y-axes.

In the language of catastrophe theory, the formula (8.19) is a (non-smooth) *unfolding* of the nongeneric caustic emanating from the nascent cusp. It also resembles a thermodynamic equation of state in the vicinity of a phase transition. However, the thermodynamic interpretation of the variables differs from the case of an on-axis focus, as analysed in the last section. Here $|y|$ is analogous to $T_c - T$, for example. And by examination, the thermodynamic critical exponent γ is nonzero whenever $\mu < 3/2$; it equals $(3/2 - \mu)^{-1}$. In any event, the critical exponents of the nascent cusp are clearly nonclassical: they depend continuously on the parameter μ.

We caution the reader that in *arbitrary* double well models at criticality, the nonpolynomial approximation (8.18) and the nonclassical equation of state (8.19) may not necessarily describe the $p_x \to 0$ behavior of $x = x(p_x, y)$ at fixed, nonzero y. If $\mu < 3/4$, the $y^{4/3}\left[|p_x|^{(4\mu/3)-1}\,\mathrm{sgn}\, p_x\right]$ term in (8.19) would cause $x = x(p_x, y)$ to diverge as $p_x \to 0$, at any nonzero y or p_y. Such a divergence would greatly distort the shape of the Lagrangian manifold $\mathcal{M}$. So we shall assume $\mu > 3/4$ henceforth. In Section 8.2 we present numerical evidence of the need for the $\mu > 3/4$ restriction, and also verify that the nongeneric caustic is present if and only if $\mu < 3/2$. Incidentally, our numerical results indicate that at fixed nonzero y, the apparent non-analyticity at $p_x = 0$ is 'rounded' at a sufficiently small ($|y|$-dependent) lengthscale, as $p_x \to 0$. This is analogous to the abovementioned rounding of $\widetilde{W}^{(y)}(x, p_y)$, and its derivative $y = y(x, p_y)$, as $p_y \to 0$ at fixed nonzero x.

There is also a problem with the nonpolynomial approximation (8.18) and the nonclassical equation of state (8.19) when $\mu \geq 3$. To see this, note that a more complete asymptotic expansion of $x = x(p_x, y)$ near $p_x = 0$ would presumably be of the form

$$x = x(p_x, y) \approx \left(-p_x/2\lambda_x + \cdots + \text{const} \times p_x^{\lfloor 4\mu \rfloor -1}\right) + c_{x;p_x,y} y^{4/3} \left[|p_x|^{4\mu/3-1} \operatorname{sgn} p_x\right]. \quad (8.21)$$

Such an asymptotic expansion is listed in Fig. 10, and is certainly valid as $\boldsymbol{x} \to \boldsymbol{0}$ from within Region N. If a similar expansion is valid near the y-axis, we see that there will be a crossover at $\mu = 3$ between two regimes. When $\mu = 3$, the nonpolynomial $c_{x;p_x,y} y^{4/3} \left[|p_x|^{4\mu/3-1} \operatorname{sgn} p_x\right]$ term in (8.21) becomes $c_{x,p_x,y} y^{4/3} p_x^3$. This is increasingly dominated by the p_x^3 term in (8.21) as $y \to 0$. In fact when μ is raised above 3, at small $|y|$ the leading corrections to the naive $x \sim -p_x/2\lambda_x$ behavior are no longer given by the nonpolynomial term, but rather by the p_x^3 term. For this reason we shall assume for the remainder of our analysis that $\mu < 3$ as well as $\mu > 3/4$.

To use the one-dimensional Maslov-WKB diffraction integral (6.5) as promised, we need to approximate in the vicinity of the nascent cusp at $(p_x, y) = (0, 0)$ not only the Legendre-transformed action $\widetilde{W}^{(x)}(p_x, y)$, but also the transformed prefactor $\widetilde{K}^{(x)}(p_x, y)$. It may be approximated in a very similar way, which we only summarize. By (5.54)–(5.55), $K(x, 0)$ diverges in any critical model as $|x|^{-\mu}$, when $x \to 0$. A scaling form

$$K(x, y) \sim |x|^{-\mu} q(y/|x|^{2\mu}), \quad (8.22)$$

modelled after (7.16) and (8.4), may be used to approximate K away from the x-axis. The approximation (8.22) should be accurate to $O(|x|^{-\mu})$ as $x \to 0$, when $y = O(|x|^{2\mu})$ [*i.e.*, as $\boldsymbol{x} \to \boldsymbol{0}$ from within Region N]. By substituting (8.22) and the scaling form (8.4) for W into the amplitude transport equation (5.7), and working to leading order near $(x, y) = (0, 0)$, one can determine the scaling function $q = q(z)$. The procedure closely resembles the procedure used in Section 7. It is easily verified that collecting the $O(|x|^{-\mu})$ terms in the transport equation yields the ODE

$$2\Big[h' + |\lambda_y|\, z\Big] q' + h'' q = 0, \quad (8.23)$$

which $q = q(z)$ must satisfy. (Cf. (7.17).) Here $h = h(z)$ is the scaling function for W, which satisfies the ODE (8.6). Using elementary calculus, one can show that eq. (8.23) has solution

$$q(z) = \text{const} \times |h'(z)|^{-1/3} \sqrt{-h''(z)}. \quad (8.24)$$

(Cf. (7.18).) But since $p_y = \partial W/\partial y \sim |x|^{2\mu} h'$, one may write $|x|^{-2\mu} p_y$ for h', and $\partial p_y/\partial y$ for h''. Substituting (8.24) into the scaling form (8.22), and performing the indicated rewriting, yields

$$K(x, y) \sim \text{const} \times |x|^{-\mu/3} |p_y|^{-1/3} \sqrt{-\frac{\partial p_y}{\partial y}(x, y)}. \quad (8.25)$$

(Cf. (7.20).) This asymptotic approximation is exact to leading order as $\boldsymbol{x} \to \boldsymbol{0}$ from within Region N.

The formula (8.25) facilitates the computation of the transformed prefactor $\widetilde{K}^{(y)} = \widetilde{K}^{(y)}(x, p_y)$. It may be computed from K as in (7.21), by dividing by the appropriate 'Van Vleck factor.' We immediately find

$$\widetilde{K}^{(y)}(x, p_y) \propto K(x, y) \bigg/ \sqrt{-\frac{\partial^2 W}{\partial y^2}(x, y)} \tag{8.26}$$

$$\sim \text{const} \times |x|^{-\mu/3} |p_y|^{-1/3}, \tag{8.27}$$

since $\partial^2 W/\partial y^2 = \partial p_y/\partial y$. (Cf. (7.21)–(7.22).) The uncomplicated asymptotic approximation (8.27) should be accurate to $O(|x|^{-\mu})$ as $x \to 0$, when $p_y = O(|x|^{2\mu})$. It simply says that $\widetilde{K}^{(y)}(x, a|x|^{2\mu}) \sim \text{const} \times |a|^{-1/3}|x|^{-\mu}$ as $x \to 0$, for any nonzero a.

The two remaining transformed prefactors, $\widetilde{K}$ and $\widetilde{K}^{(x)}$, may be computed from $\widetilde{K}^{(y)}$ by dividing (or multiplying) by the appropriate Van Vleck factors. (Cf. (6.4) and (6.6).) For example,

$$\widetilde{K}(p_x, p_y) \propto \widetilde{K}^{(y)}(x, p_y) \sqrt{-\frac{\partial^2 \widetilde{W}}{\partial p_x^2}(p_x, p_y)}. \tag{8.28}$$

The details are left to the reader. One finds

$$\widetilde{K}(p_x, p_y) \sim \text{const} \times |p_x|^{-\mu/3} |p_y|^{-1/3}, \tag{8.29}$$

$$\widetilde{K}^{(x)}(p_x, y) \sim \text{const} \times |p_x|^{-\mu/3} |y|^{-1/3}. \tag{8.30}$$

The transformed prefactor $\widetilde{K}^{(x)}$ is the one we need for the Maslov-WKB diffraction integral. In Section 8.2, we examine the numerical evidence for the validity of this asymptotic approximation to $\widetilde{K}^{(x)} = \widetilde{K}^{(x)}(p_x, y)$, when $p_x \to 0$ at fixed nonzero y.

Remarkably, the formula (8.30) predicts that $\widetilde{K}^{(x)}(p_x, y)$ *diverges* at the location $(p_x, y) = (0, 0)$ of the nascent cusp. This is different from the case of a generic (cusp) singularity, treated in Section 7. It is also different from the geometrical optics limit of physical optics, where the transformed amplitude function near a singularity is normally a 'slowly varying' (*i.e.*, non-singular) function [4]. The fact that the transformed prefactor $\widetilde{K}^{(x)}$ diverges at the nascent cusp is at least as important to the weak-noise behavior of critical double well models as the fact that the normal form for the transformed action $\widetilde{W}^{(x)}$ is nonpolynomial.

8.2 Comparison with Numerics

We now summarize the numerical evidence for the validity, in double well models at criticality, of our nonpolynomial normal form for the Legendre-transformed action $\widetilde{W}^{(x)} = \widetilde{W}^{(x)}(p_x, y)$, and our approximation to the transformed WKB prefactor $\widetilde{K}^{(x)} = \widetilde{K}^{(x)}(p_x, y)$. We shall see that both are valid approximations near the y-axis separatrix; in particular, near the saddle point. This justifies their use in the Maslov-WKB method, which we shall employ in Section 8.3 to construct boundary layer approximations to the stationary and quasistationary probability distributions of double well models at criticality.

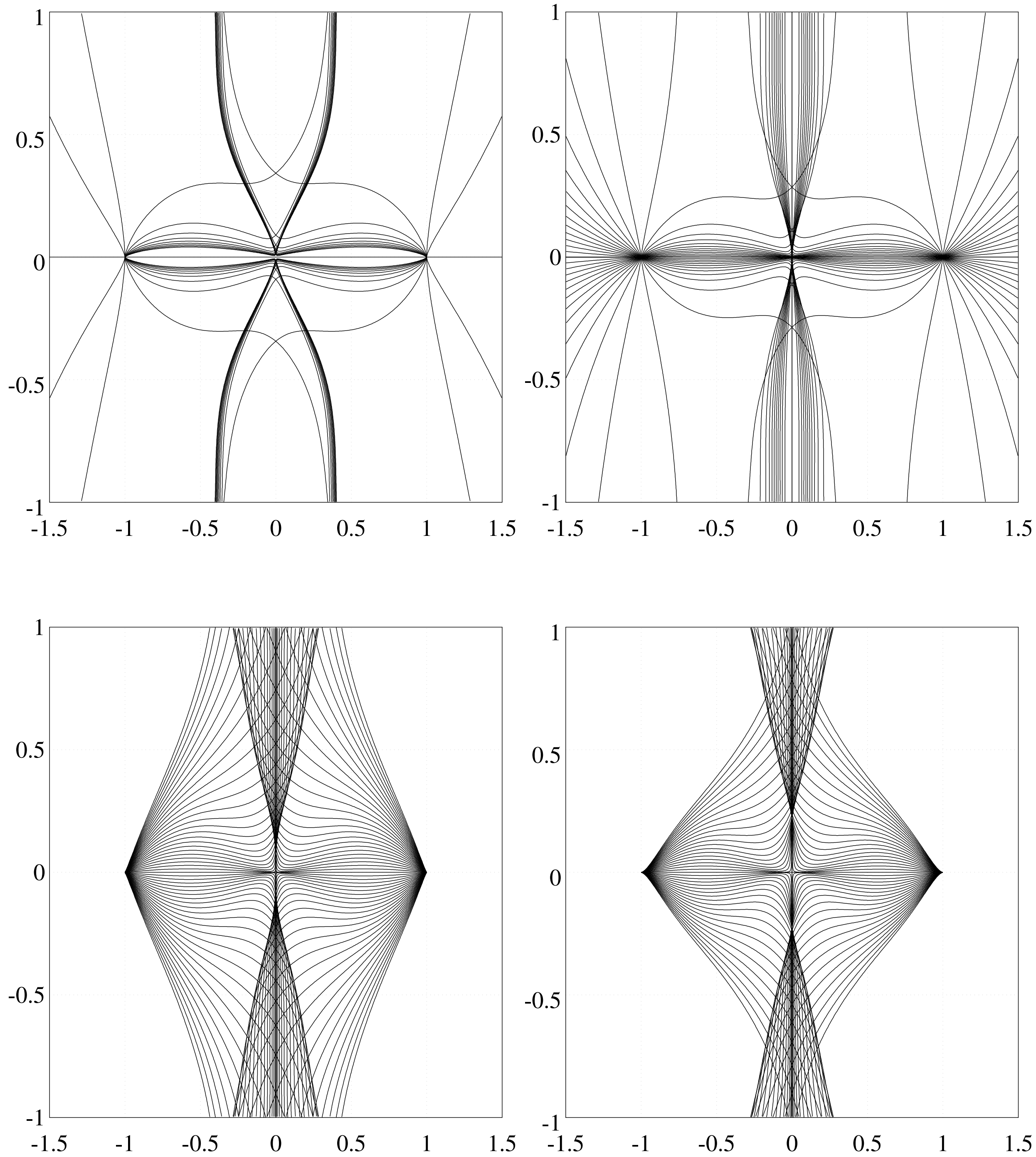

Figure 11: The flow field of instanton trajectories emanating from both stable fixed points S and S', in critical versions of the standard double well model (3.1). Parts (a), (b), (c), (d) correspond to models with $\mu = 0.725$, 0.85, 1.15, and 1.6. In all cases the parameter α is set equal to the critical value $\alpha_c = 2\mu(\mu + 1)$, at which the MPEP bifurcates. The two-sided nongeneric caustic of Fig. 4 is visible in parts (b) and (c), but it has separated into two generic caustics in part (d).

We begin by examining the evidence for the nonpolynomial normal form (8.18) for $\widetilde{W}^{(x)} = \widetilde{W}^{(x)}(p_x, y)$. Actually we shall study the related nonpolynomial approximation (8.19) to its first derivative $x = x(p_x, y)$, *i.e.*,

$$x = x(p_x, y) \approx -p_x/2\lambda_x + c_{x;p_x,y}\, y^{4/3}\Big[|p_x|^{(4\mu/3)-1}\,\mathrm{sgn}\, p_x\Big]. \tag{8.31}$$

As explained above, we expect on theoretical grounds that this approximation is generically valid near the saddle point, in critical models in which the quotient $\mu \equiv |\lambda_y|\,/\lambda_x$ satisfies $3/4 < \mu < 3$. The formula (8.31) predicts that at nonzero y, the correspondence $p_x \mapsto x$ is monotone if $\mu \geq 3/2$, but non-monotone at nonzero y if $\mu < 3/2$. When $\mu < 3/2$, the correspondence $p_x \mapsto x$ is analogous to the correspondence $\mathfrak{m} \mapsto -\mathfrak{h}$, in a ferromagnet, between magnetization and (negative) magnetic field.

It is easily checked that when

$$|x| \lesssim \text{const} \times |y|^{(3/2-\mu)^{-1}}, \quad y \to 0, \tag{8.32}$$

the inverse map $x \mapsto p_x$ is three-valued rather than single-valued. In this region the three possible values for p_x are by examination of the same magnitude as x, *i.e.*,

$$p_x = O\left(|y|^{(3/2-\mu)^{-1}}\right), \quad y \to 0. \tag{8.33}$$

We interpret the inequality (8.32) as defining a two-sided nongeneric caustic centered on the y-axis, in the interior of which the action W, and its gradient $\boldsymbol{p} = \boldsymbol{\nabla} W$, are three-valued. Recall that the Lagrangian manifold $\mathcal{M}$, which is traced out by WKB bicharacteristics, comprises all points in phase space of the form $(\boldsymbol{x}, \boldsymbol{p}(\boldsymbol{x}))$. The three-valuedness of W and $\boldsymbol{p}$ within the caustic accordingly implies that there are three points on $\mathcal{M}$ 'above' any point $\boldsymbol{x}$ in the interior of the caustic. We have already seen in Fig. 4 that a nongeneric caustic qualitatively agreeing with this prediction does indeed appear in the $\mu = 1$ standard double well model (3.1), at criticality.

Figures 11(a)–11(d) show the flow field of instanton trajectories, *i.e.*, projected bicharacteristics, in several more critical variants of the standard double well model. (At criticality $\alpha = \alpha_c = 2\mu(\mu + 1)$, by eq. (5.50).) Figures 11(b) and 11(c), with $\mu = 0.85$ and $\mu = 1.15$, illustrate the fact that the two-sided caustic of Fig. 4 appears at criticality in any double well model whose parameter $\mu \equiv |\lambda_y|/\lambda_x$ satisfies $3/4 < \mu < 3/2$. The caustic disappears, as expected, in critical models with $\mu \geq 3/2$. Figure 11(d) shows what happens. As μ in the standard model is raised above 3/2 (with α set equal to $\alpha_c = \alpha_c(\mu)$), the two-sided caustic separates into two one-sided *generic* caustics, whose cusps move out along the positive and negative y-axes, away from the saddle. In any critical model with $\mu > 3/2$, there is a portion of the separatrix near the saddle that is not crossed by any instanton trajectory.

Figure 11(a) illustrates the bizarre behavior that occurs in critical models with $\mu \leq 3/4$. At first glance it seems that the now familiar two-sided caustic is present, but closer study reveals that points in its interior are reached by only *two* instanton trajectories, rather than three. Apparently, in the $\mu \leq 3/4$ regime the approximation (8.31) breaks down near the separatrix. Empirically, when $\mu \leq 3/4$ the $|p_x|^{(4\mu/3)-1}$ factor in (8.31) must be replaced by unity. The $\mu \leq 3/4$ regime is still under investigation, and we shall not consider it further in this paper.

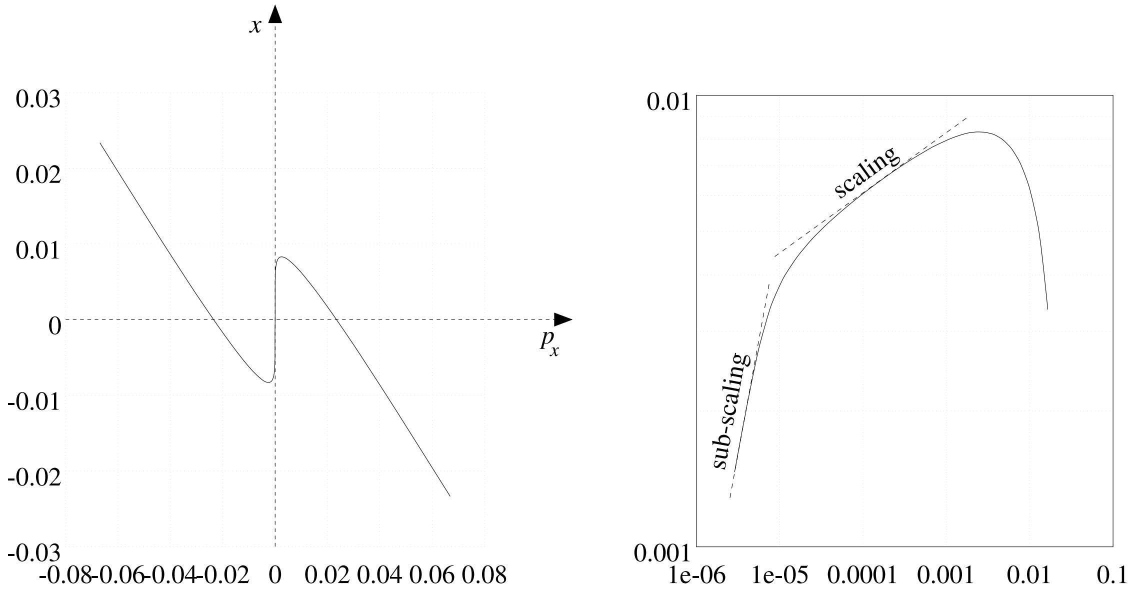

Figure 12: A section through the nongeneric caustic appearing at criticality in the $\mu = 0.85$ standard double well model, as shown in Fig. 11(b). This section is taken at $y = 0.05$. Parts (a) and (b) show a linear and a logarithmic plot of $x = x(p_x, y = 0.05)$. The scaling behavior ($x \propto |p_x|^{(4\mu/3)-1}$) and the sub-scaling behavior ($x \propto p_x$) are both visible. Crossover between the two regimes occurs at $p_x = O(|y|^{(5/2)(3/2-\mu)^{-1}})$, *i.e.*, at $p_x \doteq 10^{-5}$.

We can now compare the predictions of our scaling theory with numerical data. Figure 12(a) is a section through the caustic of Fig. 11(b), *i.e.*, a cross-section through the corresponding Lagrangian manifold. It shows the correspondence $p_x \mapsto x$, at $y = 0.05$, in the $\mu = 0.85$ standard model at criticality. The qualitative shape of the curve certainly resembles the prediction of formula (8.31). But before making a *quantitative* comparison, we need to discuss the interpretation of (8.31). It was derived from an asymptotic development of the action about the saddle point. To what extent does it describe the small-$|p_x|$ asymptotics of $x = x(p_x, y)$ at *fixed*, nonzero y? That is what is plotted in Fig. 12.

From a rigorous point of view, when $\mu < 3/2$ the formula (8.31) provides a two-term asymptotic expansion of $x = x(p_x, y)$ as $y \to 0$, on the $p_x = O(|y|^{(3/2-\mu)^{-1}})$ lengthscale on which the nongeneric caustic is visible. This is strongly reminiscent of the 'Region N' constraint of the last section. There we began by approximating $W = W(x, y)$ in the notch-shaped region of Fig. 8. Here we are approximating $x = x(p_x, y)$, and by extension its anti-derivative $\widetilde{W}^{(x)} = \widetilde{W}^{(x)}(p_x, y)$, in a region that is similarly notch-shaped, but is centered on the y-axis rather than the x-axis. We shall not attempt to expand x and $\widetilde{W}^{(x)}$ systematically, but the basic procedure is plain. If we define a new scaling variable

$$\hat{z} \equiv p_x / \, |y|^{(3/2-\mu)^{-1}}, \tag{8.34}$$

then formula (8.31) can be interpreted as comprising the first two terms in an asymptotic development of $x(\hat{z}|y|^{(3/2-\mu)^{-1}}, y)$ as $y \to 0$. The development should be valid at any fixed $\hat{z}$.

Though this restatement is a bit pedantic, it suggests that on smaller lengthscales than $p_x = O(|y|^{(3/2-\mu)^{-1}})$, the formula (8.31) might be invalid. Actually there are strong reasons for believing that the nonanalytic $|p_x|^{(4\mu/3)-1}\operatorname{sgn} p_x$ behavior as $p_x \to 0$ does not appear at fixed, nonzero y. If it did, $\partial x/\partial p_x(p_x = 0)$ would diverge at criticality, at $y \neq 0$, in any model with $3/4 < \mu < 3/2$. Equivalently, $\partial p_x/\partial x(x = 0)$, *i.e.*, $\partial^2 W/\partial x^2(x = 0)$, would be *identically zero*, irrespective of the choice of nonzero y. But this prediction is too simple: it ignores the presence of 'sub-scaling' terms. We noted in the last section that at criticality, $W = W(x, y)$ should contain an x^2y^2 term, for consistency with the sub-scaling $w_2(x) \sim \text{const} \times x^2$ behavior. In other words, $\partial^2 W/\partial x^2(x = 0)$ near the saddle should be nonzero and proportional to y^2. For consistency with this prediction, the nonanalytic $|p_x|^{(4\mu/3)-1}\operatorname{sgn} p_x$ behavior of $x = x(p_x, y)$ must be 'rounded' at sufficiently small $|p_x|$. It is easy to check that $p_x = O(|y|^{(5/2)(3/2-\mu)^{-1}})$ is the correct lengthscale. On that lengthscale, one should find $W(x, y) \approx \text{const} \times x^2y^2$, *i.e.*, $p_x \approx \text{const} \times xy^2$, or $x \approx \text{const} \times y^{-2}p_x$.

What we conclude from this discussion is that at fixed, nonzero y, the nonpolynomial formula (8.31) for $x = x(p_x, y)$ should be valid on the 'caustic lengthscale' $p_x = O(|y|^{(3/2-\mu)^{-1}})$, but that it will break down when p_x is decreased to $O(|y|^{(5/2)(3/2-\mu)^{-1}})$. On that smaller lengthscale, one expects a crossover to a linear regime, where x is proportional to p_x. We can now proceed to our comparison with numerics. In Fig. 12(b) we plot the correspondence $p_x \mapsto x(p_x, y = 0.05)$ of Fig. 12(a) on a logarithmic scale. We also fit two trendlines to it: $x \propto |p_x|^{(4\mu/3)-1}$ and $x \propto p_x$. As the two trendlines reveal, our theoretical analysis is perfectly confirmed. There is indeed a crossover to a linear, sub-scaling regime when p_x is decreased to $O(|y|^{(5/2)(3/2-\mu)^{-1}})$. But at larger lengthscales, *e.g.*, $p_x = O(|y|^{(3/2-\mu)^{-1}})$, the fractional power $|p_x|^{(4\mu/3)-1}\operatorname{sgn} p_x$ of the scaling formula (8.31) is clearly visible. Similar crossover plots can be obtained for other critical double well models, whose parameter $\mu = |\lambda_y|/\lambda_x$ lies in the range $3/4 < \mu < 3/2$.

Our asymptotic approximation $\widetilde{K}^{(x)}(p_x, y) \sim \text{const} \times |p_x|^{-\mu/3}|y|^{-1/3}$ to the transformed prefactor $\widetilde{K}^{(x)}$ at criticality, derived in Section 8.1, can also be numerically tested. The approximation should be valid in the vicinity of the y-axis separatrix, *i.e.*, as $p_x \to 0$ at fixed, nonzero y. There are two separate cases: $3/4 < \mu < 3/2$, when a caustic is present, and $\mu \geq 3/2$, when one is not. For simplicity we consider only the latter. When $\mu \geq 3/2$, it follows from (8.31) that $x(p_x, y) \approx -p_x/2\lambda_x$ as $p_x \to 0$; the nonpolynomial term is subdominant. So our asymptotic approximation to $\widetilde{K}^{(x)}$ implies that (cf. (6.6))

$$K(x, y) \propto \widetilde{K}^{(x)}(p_x, y)\sqrt{-\frac{\partial^2 W}{\partial x^2}(p_x, y)} \tag{8.35}$$

$$= \widetilde{K}^{(x)}(p_x, y)\sqrt{-\frac{\partial p_x}{\partial x}(p_x, y)} \tag{8.36}$$

$$\approx \text{const} \times |x|^{-\mu/3}\,|y|^{-1/3}, \tag{8.37}$$

as $x \to 0$ at fixed, nonzero y. This comparatively slow power-law divergence as the separatrix is approached at (small) nonzero y is to be contrasted with the $K(x, 0) \sim \text{const} \times |x|^{-\mu}$ divergence that occurs when the saddle point is approached along the x-axis. (See (5.55).) It is susceptible to numerical test.

In Fig. 13 we graph $K = K(x, y)$ as a function of x, at $y = 0.05$, for the critical version of the

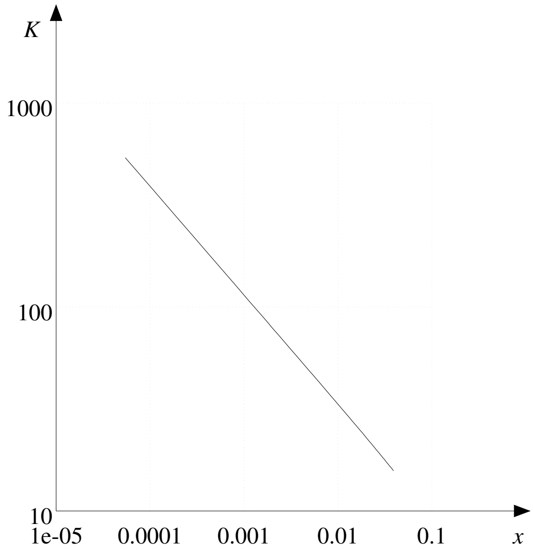

Figure 13: Behavior of the WKB prefactor near the y-axis separatrix, in a critical version of the $\mu = 1.6$ standard double well model. $K = K(x, y = 0.05)$ is plotted on a logarithmic scale, revealing the scaling behavior ($K \propto x^{-\mu/3}$ at nonzero y).

standard double well model with $\mu = 1.6$. (This is the same model whose instanton trajectories are shown in Fig. 11(d).) The curve is fitted to high accuracy by a power-law const $\times\, x^{-0.533}$, *i.e.*, const $\times\, x^{-\mu/3}$. This confirms the prediction of our scaling theory. No sub-scaling regime is evident at small $|x|$. Similar plots can be obtained for the near-separatrix behavior of K in critical models with other values of μ.

We conclude that the asymptotic approximations to $\widetilde{W}^{(x)} = \widetilde{W}^{(x)}(p_x, y)$ and $\widetilde{K}^{(x)} = \widetilde{K}^{(x)}(p_x, y)$ derived from our scaling theory have a wide domain of validity, and may be employed in the Maslov-WKB method.

8.3 Scaling Beyond the WKB Approximation

We can now compute the Maslov-WKB boundary layer approximations to the stationary density ρ_0 and the quasistationary density ρ_1 near the nascent cusp at the saddle point, where the conventional WKB approximation breaks down. The boundary layer approximations are determined by (8.18) and (8.30), the asymptotic approximations to the Legendre-transformed action $\widetilde{W}^{(x)}$ and the transformed prefactor $\widetilde{K}^{(x)}$ respectively. As discussed above, these approximations should be valid when the eigenvalue ratio $\mu \equiv |\lambda_y|/\lambda_x$ satisfies $3/4 < \mu < 3$.

Substituting (8.18) and (8.30) into the one-dimensional diffraction integral (6.5) yields the rather complicated expression

$$
\begin{aligned}
&K(x,y)\exp[-W(x,y)/\epsilon] \qquad (8.38)\\
&\quad\sim \epsilon^{-1/2}\int \widetilde{K}^{(x)}(p_x,y)\exp\left\{\left[-xp_x+\widetilde{W}^{(x)}(p_x,y)\right]/\epsilon\right\}\,dp_x\\
&\quad\approx \text{const }\times \epsilon^{-1/2}e^{-W(0,0)/\epsilon}|y|^{-1/3}e^{-|\lambda_y|y^2/\epsilon}\\
&\qquad\qquad\times\int |p_x|^{-\mu/3}\exp\left\{-\left[\frac{p_x^2}{4\lambda_x}-C_{p_x,y}y^{4/3}|p_x|^{4\mu/3}+xp_x\right]\Big/\,\epsilon\right\}\,dp_x
\end{aligned}
$$

which requires a bit of explanation. The first problem to be resolved is the lengthscale near $(x, y) = (0, 0)$ on which this diffraction integral defines a valid Maslov-WKB approximation. The p_x^2 and xp_x terms in the argument of $\exp(\cdot)$ are $O(1)$ when p_x^2 and xp_x are $O(\epsilon)$; *i.e.*, when x and p_x are $O(\epsilon^{1/2})$. The term $C_{p_x,y}y^{4/3}|p_x|^{4\mu/3}/\epsilon$ is $O(1)$ when, also, $y = O(\epsilon^{3/4-\mu/2})$. This is the case when $\mu < 3/2$, at least. If $\mu \geq 3/2$ then the $C_{p_x,y}y^{4/3}|p_x|^{4\mu/3}/\epsilon$ term is negligible whenever $y = o(1)$. We conclude that in the weak-noise ($\epsilon \to 0$) limit of models with $\mu < 3/2$, the diffraction integral (8.38) defines a valid Maslov-WKB approximation on the $x = O(\epsilon^{1/2})$, $y = O(\epsilon^{3/4-\mu/2})$ lengthscale near the saddle point. This is precisely the caustic lengthscale $x = O(|y|^{(3/2-\mu)^{-1}})$, *i.e.*, $p_x = O(|y|^{(3/2-\mu)^{-1}})$, of the last section. If $\mu \geq 3/2$, so that no caustic is present, then $y = O(\epsilon^{3/4-\mu/2})$ must be replaced by $y = o(1)$. On the appropriate lengthscale, the diffraction integral defines a *noncanonical* diffraction function.

The diffraction integral, being one-dimensional, cannot resolve the singularity at $y = 0$. This is because it does not include an integration over p_y. So one cannot expect the Maslov-WKB approximation to be valid at arbitrarily small y. The stationary and quasistationary densities in critical models are expected to be tightly concentrated on the $y = O(\epsilon^{1/2})$ transverse lengthscale near the saddle point, as we saw in (5.19)–(5.21) (which apply in the absence of focusing). At most, the Maslov-WKB approximation will be valid in the *far field* of the $y = O(\epsilon^{1/2})$ lengthscale, where the factor $e^{-|\lambda_y|y^2/\epsilon}$ is exponentially small. So it will not be directly comparable to (5.19)–(5.21). But it proves to be very useful nonetheless. Define 'stretched' variables $X \equiv x/\epsilon^{1/2}$ and $Y \equiv y/\epsilon^{1/2}$; also, change the integration variable to $P_x = p_x/\epsilon^{1/2}$. The approximation becomes

$$
\text{const} \times \epsilon^{-(\mu+1)/6}e^{-W(0,0)/\epsilon}\,|Y|^{-1/3}\,e^{-|\lambda_y|Y^2}\int |P_x|^{-\mu/3}\,e^{-XP_x}e^{-P_x^2/4\lambda_x}\,dP_x, \qquad (8.39)
$$

since when $p_x, y = O(\epsilon^{1/2})$, the term $C_{p_x,y}y^{4/3}\,|p_x|^{4\mu/3}/\epsilon$ is negligible.

At this point we must explain how to interpret the integration over P_x, or p_x. The stationary density $\rho_0(x, y) = \rho_0(X\epsilon^{1/2}, Y\epsilon^{1/2})$ must be even in X, and the quasistationary density $\rho_1(x, y) = \rho_1(X\epsilon^{1/2}, Y\epsilon^{1/2})$ must be odd. To get approximations with these symmetry properties, we may integrate P_x from $-\infty$ to ∞ and 0 to ∞ respectively. We remarked at the end of Section 6 that performing a half-range integration is one way of incorporating an antisymmetry constraint, and that is the technique we shall use.

As summarized in Abramowitz and Stegun (Ref. [1], §19.5), definite integrals resembling (8.39) define *parabolic cylinder functions*. Evaluating the integral (8.39), with the two possible choices of the range of integration (full range and half range), yields the Maslov-WKB approximations

$$\rho_0(x,y) \sim \text{const} \times \epsilon^{-(\mu+1)/6} F_0(\lambda_x^{1/2} x/\epsilon^{1/2}) e^{+\lambda_x x^2/\epsilon} \left| y/\epsilon^{1/2} \right|^{-1/3} e^{-|\lambda_y| y^2/\epsilon}, \tag{8.40}$$

$$\rho_1(x,y) \sim \text{const} \times \epsilon^{-(\mu+1)/6} F_1(\lambda_x^{1/2} x/\epsilon^{1/2}) e^{+\lambda_x x^2/\epsilon} \left| y/\epsilon^{1/2} \right|^{-1/3} e^{-|\lambda_y| y^2/\epsilon} \tag{8.41}$$

to the stationary and quasistationary probability densities, on the $O(\epsilon^{1/2})$ lengthscale near the saddle point. Here the so-called boundary layer functions $F_i = F_i(Z)$, where $Z \equiv \lambda_x^{1/2} X = \lambda_x^{1/2} x/\epsilon^{1/2}$, are defined by

$$F_i(Z) \equiv y_{i+1}(1/2 - \mu/3, 2^{1/2} Z) e^{-Z^2/2}, \tag{8.42}$$

in the notation of Abramowitz and Stegun. $y_1(1/2 - \mu/3, \bullet)$ and $y_2(1/2 - \mu/3, \bullet)$ are even and odd parabolic cylinder functions, respectively. We could equally well define the boundary layer functions F_i in terms of an *Hermite function of non-integer index*, by

$$F_0(Z) \equiv \left[H_{(\mu/3)-1}(Z) e^{-Z^2} \right]_{\text{even}}, \qquad F_1(Z) \equiv \left[H_{(\mu/3)-1}(Z) e^{-Z^2} \right]_{\text{odd}}. \tag{8.43}$$

Here $[\cdot]_{\text{even}}$ and $[\cdot]_{\text{odd}}$ signify even and odd parts, under the reflection $Z \mapsto -Z$. The definitions (8.43) are meaningful whenever the index $n \equiv (\mu/3) - 1$ is not an integer, in which case the Hermite function $H_n(Z)$ is not a conventional Hermite polynomial, and is neither even nor odd. But since we are assuming $3/4 < \mu < 3$, this is always the case. Irrespective of the choice of definitions,

$$\begin{aligned} F_0(Z) &\sim \text{const} \times |Z|^{-\mu/3}, \\ F_1(Z) &\sim \text{const} \times |Z|^{-\mu/3} \operatorname{sgn} Z, \end{aligned} \tag{8.44}$$

as $Z \to \pm\infty$. (Cf. Ref. [1], §19.8.)

We stress that the Maslov-WKB approximations to ρ_0 and ρ_1 are strictly valid only in the *transverse far field*, *i.e.*, as $Y = y/\epsilon^{1/2} \to \pm\infty$. But they make very clear how critical double well models differ from non-critical double well models. By comparing (8.40)–(8.41) with (5.19)–(5.21), we see that at criticality, the boundary layer functions $F_0(\cdot)$ and $F_1(\cdot)$ replace the boundary layer functions 1 and $\operatorname{erf}(\cdot)$ respectively. The approximations (8.40)–(8.41) are guaranteed to match to the standard WKB approximation $K(\boldsymbol{x}) \exp[-W(\boldsymbol{x})]$, as one moves in a transverse direction away from the saddle point. For example, the $|Z|^{-\mu/3}$ falloff of eqs. (8.44) will match to the $|x|^{-\mu/3}$ prefactor falloff of eqs. (8.25) and (8.27), which is seen, *e.g.*, in Figure 13.

The Maslov-WKB approximations to ρ_0 and ρ_1, and the nonpolynomial normal form for the Legendre-transformed action that engendered them, have several striking consequences for double well models at criticality.

- A nongeneric caustic, emerging sideways from the nascent cusp at the saddle. In Section 8.1 we predicted from the nonpolynomial normal form for $\widetilde{W}$ that when $3/4 < \mu < 3/2$, a caustic is located at $|x| \lesssim \text{const} \times |y|^{(3/2-\mu)^{-1}}$. Our prediction was confirmed by Figure 11. This caustic has an unusual (continuously varying) exponent. It is *nongeneric*, in the sense of singularity theory.

- An unusual (continuously varying) singularity index. As $\epsilon \to 0$, the falloff of the stationary density ρ_0 at the saddle point $(0, 0)$ is not pure exponential, on account of the $\epsilon^{-(\mu+1)/6}$ prefactor in the Maslov-WKB approximation (8.40). This is interpreted as a statement that the nascent cusp has singularity index $s = s(\mu) = (\mu + 1)/6$, as mentioned in Section 4. It too is nongeneric, in the sense of singularity theory.

- Non-Arrhenius MFPT asymptotics. If one computes the rate at which the quasistationary density ρ_1 is absorbed on the separatrix near the saddle, the $\epsilon^{-(\mu+1)/6}$ prefactor in the Maslov-WKB approximation (8.41) will appear in the $\epsilon \to 0$ asymptotics. Equivalently, the exponentially decaying eigenvalue $\lambda_1 = \lambda_1(\epsilon)$ of the Smoluchowski operator will have an asymptotic $\epsilon^{-(\mu+1)/6}$ prefactor, as well as the usual Arrhenius factor [*i.e.*, $\exp(-\Delta W/\epsilon)$]. And the MFPT will be asymptotic to const $\times\ \epsilon^{+(\mu+1)/6} \exp(+\Delta W/\epsilon)$, as $\epsilon \to 0$. At criticality, the weak-noise growth of the MFPT is *slower than pure exponential*.

- A non-Gaussian limiting exit location distribution. In the absence of MPEP bifurcation, for a symmetric double well model the location of the point of exit from either of the two wells would have an asymptotic Gaussian distribution, on the transverse $O(\epsilon^{1/2})$ lengthscale near the saddle. In fact, its density would fall off as $e^{-|\lambda_y|y^2/\epsilon}$. We see from the Maslov-WKB approximation to ρ_1 that at criticality, the exit location density on the separatrix includes *scaling corrections*. In the transverse far field it falls off as $|y|^{-1/3}e^{-|\lambda_y|y^2/\epsilon}$, rather than $e^{-|\lambda_y|y^2/\epsilon}$. It has a *non-Gaussian tail*.

These phenomena look natural from the point of view of the theory of critical phenomena, though the stochastic escape problem has not previously been considered from that point of view.

9 Discussion

We can now step back and review our results. We began with a WKB treatment of the weak-noise asymptotics of stationary (and quasistationary) solutions of the Smoluchowski equation. The WKB analysis led to instanton trajectories, which have a physical interpretation as most probable weak-noise fluctuational paths. The instanton trajectories turned out to be zero-energy trajectories of an associated Hamiltonian dynamical system. This is because the phase space versions of the instanton trajectories (*i.e.*, WKB bicharacteristics) trace out a Lagrangian manifold in phase space. In double well models the onset of bifurcation is associated with the fleeting appearance of an unusual singularity (a nascent cusp) in the shape of this manifold, as the parameters of the model are varied.

There is a formal analogy between the Lagrangian manifold of a dynamical system perturbed by weak noise, and the thermodynamic surface of a condensed matter system. This analogy led us to construct a scaling theory of the shape of the Lagrangian manifold near the nascent cusp. To date, most work on Lagrangian manifolds has assumed that they are *smooth*, and that any apparent singularities in their shape can be transformed away by a change of coordinates. This is analogous to assuming that thermodynamic surfaces are real analytic, and that non-analyticities in thermodynamic behavior (*i.e.*, phase transitions) can be transformed away by working in terms of

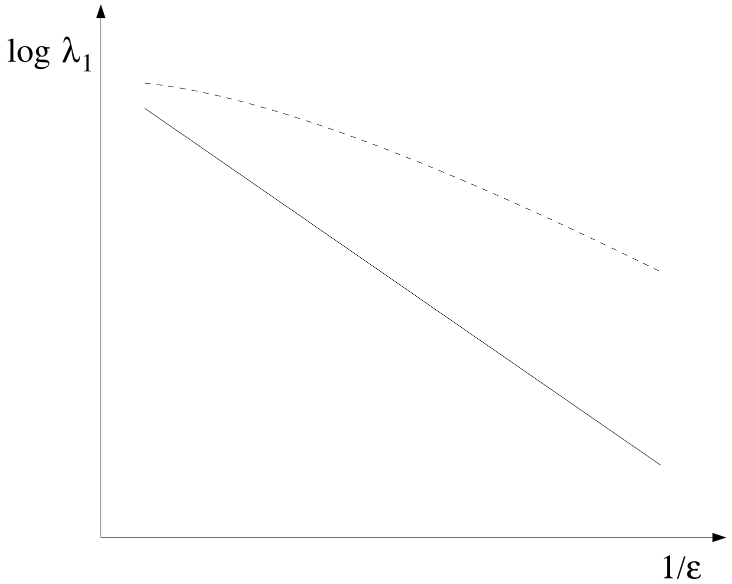

Figure 14: A sketch, on a logarithmic scale, of the rate of noise-activated inter-well hopping λ_1 as a function of the reciprocal noise strength $1/\epsilon$. Off criticality (solid curve), λ_1 displays a pure exponential falloff as $\epsilon \to 0$, *i.e.*, $\lambda_1 \sim \text{const} \times \exp(-\Delta W/\epsilon)$. At the onset of bifurcation (dashed curve), $\lambda_1 \sim \text{const} \times \epsilon^{-s} \exp(-\Delta W/\epsilon)$ instead, where s is the singularity index of the nascent cusp appearing at the saddle point.

the appropriate thermodynamic potential. Equivalently, it is analogous to assuming that all phase transitions have classical critical exponents. Our scaling theory makes it clear that the nascent cusp singularity is a genuine *point of non-smoothness* of the Lagrangian manifold. In thermodynamic terms, it has nonclassical, indeed continuously varying, critical exponents.

Applying the Maslov-WKB method to the nascent cusp yielded several interesting predictions, which we summed up in the four bulleted items at the end of the last section. One normally expects that in a double well system perturbed by weak noise of strength ϵ, the rate of inter-well hopping $\lambda_1 = \lambda_1(\epsilon)$ will be asymptotic to a *constant* multiple of the Arrhenius factor $\exp(-\Delta W/\epsilon)$, where ΔW is an effective barrier height. Also, one expects that the distribution of exit locations (from either well) will asymptotically become a Gaussian of $O(\epsilon^{1/2})$ standard deviation, centered on the saddle point between the two wells. The Maslov-WKB method predicts that at criticality, both these phenomena are strongly altered. In particular, the factor $\exp(-\Delta W/\epsilon)$ must be replaced by $\epsilon^{-s}\exp(-\Delta W/\epsilon)$, where $s = (\mu+1)/6$ is the singularity index of the nascent cusp. (As we noted in Section 5.3, the singularity index is a sort of critical exponent.) In Fig. 14 we sketch an Arrhenius plot, showing this anomalous (non-Arrhenius) behavior.

In mathematical terms, the nascent cusp appearing at the onset of bifurcation is a *nongeneric* singularity, *i.e.*, a singularity different from any of the now classical singularities of catastrophe theory. As shown in Fig. 4, in many double well models it induces an unusual caustic in the flow field of instanton trajectories. This caustic is itself nongeneric, in that its exponent is not equal to $3/2$. As we have seen (see, *e.g.*, Fig. 12), its presence quantitatively confirms the validity of our scaling theory. It is remarkable that such nongeneric phenomena are a *generic* feature of singly parametrized symmetric double well models.

At least as developed in this paper, our scaling theory is a scaling theory of weak-noise behavior near the nascent cusp, precisely at criticality. It would be useful to treat as well models that are nearly critical, but not exactly so. Such models should display a crossover from non-Arrhenius behavior to Arrhenius behavior at sufficiently weak noise strength. By developing a *joint scaling theory*, one of the variables in which measures the distance from criticality, it should be possible to analyse this phenomenon. We expect that it is possible to derive a 'Ginzburg criterion' [27], expressing how close to criticality any given double well model should be, for the non-Arrhenius behavior of Fig. 14 to be visible. Work on this is under way.

We briefly mention two geometric features of models 'off criticality' that cry out for a theoretical explanation. A nongeneric caustic appears near the separatrix not only at criticality, but in many non-critical models as well [12]. Also, as criticality is approached (*e.g.*, as $\alpha \to \alpha_c^-$ in the standard model of (3.1)), it frequently happens that the nascent cusp is formed by a collision of two generic cusps, which move along the separatrix toward the saddle point. These phenomena can presumably be explained by an appropriate *joint unfolding*, but that is for the future.

We close by mentioning a possible extension of a more theoretical sort. In this paper we have focused exclusively on the asymptotic solutions of the time-independent weak-noise Smoluchowski equation. There is reason to believe that nongeneric singularities resembling the nascent cusp can occur, and are perhaps even widespread, in the asymptotic solutions of other singularly perturbed elliptic partial differential equations. Most WKB treatments of singularly perturbed elliptic PDE's (see, *e.g.*, Duistermaat [10]) assume that each WKB characteristic (*i.e.*, instanton trajectory) eventually leaves any bounded region of space. This assumption is violated in the Smoluchowski equation for any double well model, since the MPEP(s) terminate on the saddle point, rather than extending to infinity. We expect that when it is violated in other PDE's, analogous nongeneric singularities in formal asymptotic solutions can occur. The nongeneric singular phenomena that we have seen in this paper may simply be representatives of a larger class.

References

[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. Dover, New York, 1965.

[2] V. I. Arnol'd. Critical Points of Smooth Functions and Their Normal Forms. *Russian Math. Surveys* **30:5** (1975), 1–75.

[3] A. Auerbach and S. Kivelson. The Path Decomposition and Multidimensional Tunneling. *Nuclear Phys. B* **257 [FS 14]** (1985), 799–858.

[4] M. V. Berry. Waves and Thom's Theorem. *Adv. in Phys.* **25** (1976), 1–26.

[5] M. V. Berry and N. L. Balazs. Evolution of Semiclassical Quantum States in Phase Space. *J. Phys. A* **12** (1979), 625–642.

[6] A. J. Bray and A. J. McKane. Instanton Calculation of the Escape Rate for Activation over a Potential Barrier Driven by Colored Noise. *Phys. Rev. Lett.* **62** (1989), 493–496.

[7] B. Carmeli, V. Mujica, and A. Nitzan. Dynamics of Multidimensional Barrier Crossing in the Overdamped Limit. *Berichte der Bunsen-Gesellschaft* **95** (1991), 319–326.

[8] B. Caroli, C. Caroli, B. Roulet, and J.-F. Gouyet. A WKB Treatment of Diffusion in a Multidimensional Bistable Potential. *J. Statist. Phys.* **22** (1980), 515–536.

[9] M. V. Day. Recent Progress on the Small Parameter Exit Problem. *Stochastics* **20** (1987), 121–150.

[10] J. J. Duistermaat. Oscillatory Integrals, Lagrange Immersions, and Unfolding of Singularities. *Comm. Pure Appl. Math.* **27** (1974), 207–281.

[11] M. I. Dykman, P. V. E. McClintock, V. N. Smelyanskiy, N. D. Stein, and N. G. Stocks. Optimal Paths and the Prehistory Problem for Large Fluctuations in Noise-Driven Systems. *Phys. Rev. Lett.* **68** (1992), 2718–2721.

[12] M. I. Dykman, M. M. Millonas, and V. N. Smelyanskiy. Observable and Hidden Singular Features of Large Fluctuations in Nonequilibrium Systems. *Phys. Lett. A* **195** (1994), 53–58.

[13] J.-P. Eckmann and R. Sénéor. The Maslov-WKB Method for the (an-)Harmonic Oscillator. *Arch. Rational Mech. Anal.* **61** (1976), 153–173.

[14] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer-Verlag, New York/Berlin, 1984.

[15] S. Glasstone, K. J. Laidler, and H. Eyring. *The Theory of Rate Processes*. McGraw-Hill, New York, 1941.

[16] R. Graham. Macroscopic Potentials, Bifurcations and Noise in Dissipative Systems. In *Theory of Continuous Fokker-Planck Systems*, edited by F. Moss and P. V. E. McClintock, volume 1 of *Noise in Nonlinear Dynamical Systems*, chapter 7, pp. 225–278. Cambridge University Press, Cambridge, 1989.

[17] M. C. Gutzwiller. Periodic Orbits and Classical Quantization Conditions. *J. Math. Phys.* **12** (1971), 343–358.

[18] M. C. Gutzwiller. *Chaos in Classical and Quantum Mechanics*, volume 1 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York/Berlin, 1990.

[19] H. R. Jauslin. Nondifferentiable Potentials for Nonequilibrium Steady States. *Physica A* **144** (1987), 179–191.

[20] J. B. Keller and S. I. Rubinow. Asymptotic Solution of Eigenvalue Problems. *Ann. Physics* **9** (1960), 24–75.

[21] M. M. Kłosek-Dygas, B. M. Hoffman, B. J. Matkowsky, A. Nitzan, M. A. Ratner, and Z. Schuss. Diffusion Theory of Multidimensional Activated Rate Processes: The Role of Anisotropy. *J. Chem. Phys.* **90** (1989), 1141–1148.

[22] H. A. Kramers. Brownian Motion on a Field of Force and the Diffusion Model of Chemical Reactions. *Physica* **7** (1940), 284–304.

[23] R. Landauer. Motion out of Noisy States. *J. Statist. Phys.* **53** (1988), 233–248.

[24] J. S. Langer. Statistical Theory of the Decay of Metastable States. *Ann. Physics* **54** (1969), 258–275.

[25] R. G. Littlejohn. The Van Vleck Formula, Maslov Theory, and Phase Space Geometry. *J. Statist. Phys.* **68** (1992), 7–50.

[26] D. Ludwig. Persistence of Dynamical Systems Under Random Perturbations. *SIAM Rev.* **17** (1975), 605–640.

[27] S.-K. Ma. *Modern Theory of Critical Phenomena*. W. A. Benjamin, Reading, Massachusetts, 1976.

[28] R. S. Maier and D. L. Stein. The Effect of Focusing and Caustics on Exit Phenomena in Systems Lacking Detailed Balance. *Phys. Rev. Lett.* **71** (1993), 1783–1786.

[29] R. S. Maier and D. L. Stein. The Escape Problem for Irreversible Systems. *Phys. Rev. E* **48** (1993), 931–938.

[30] R. S. Maier and D. L. Stein. Asymptotic Exit Location Distributions in the Stochastic Exit Problem. Submitted to *SIAM J. Appl. Math.*, 1994.

[31] M. Mangel. Barrier Transitions Driven by Fluctuations, with Applications to Evolution and Ecology. *Theoret. Population Biol.* **45** (1994), 16–40.

[32] V. P. Maslov and M. V. Fedoriuk. *Semi-Classical Approximation in Quantum Mechanics*. Reidel, Boston/Dordrecht, 1981.

[33] T. Naeh, M. M. Kłosek, B. J. Matkowsky, and Z. Schuss. A Direct Approach to the Exit Problem. *SIAM J. Appl. Math.* **50** (1990), 595–627.

[34] L. Onsager and S. Machlup. Fluctuations and Irreversible Processes. *Phys. Rev.* **91** (1953), 1505–1512.

[35] R. B. Paris. The Asymptotic Behaviour of Pearcey's Integral for Complex Variables. *Proc. Roy. Soc. London Ser. A* **432** (1991), 391–426.

[36] L. S. Schulman. *Techniques and Applications of Path Integration*. Wiley, New York, 1981.

[37] M. A. Shayman. A Geometric View of the Matrix Riccati Equation. In *The Riccati Equation*, edited by S. Bittanti, A. J. Laub, and J. C. Willems, pp. 89–112. Springer-Verlag, New York/Berlin, 1991.

[38] P. Talkner. Mean First Passage Times and the Lifetime of a Metastable State. *Z. Phys. B* **68** (1987), 201–207.