The Annals of Probability
2018, Vol. 46, No. 5, 2749–2799
https://doi.org/10.1214/17-AOP1238

CONTROLLED EQUILIBRIUM SELECTION IN STOCHASTICALLY PERTURBED DYNAMICS

BY ARI ARAPOSTATHIS[*,1] ANUP BISWAS[†,2] AND VIVEK S. BORKAR[‡,3]

University of Texas at Austin[*], Indian Institute of Science Education and Research, Pune[†] and Indian Institute of Technology Bombay[‡]

We consider a dynamical system with finitely many equilibria and perturbed by small noise, in addition to being controlled by an "expensive" control. The controlled process is optimal for an ergodic criterion with a running cost that consists of the sum of the control effort and a penalty function on the state space. We study the optimal stationary distribution of the controlled process as the variance of the noise becomes vanishingly small. It is shown that depending on the relative magnitudes of the noise variance and the "running cost" for control, one can identify three regimes, in each of which the optimal control forces the invariant distribution of the process to concentrate near equilibria that can be characterized according to the regime. We also obtain moment bounds for the optimal stationary distribution. Moreover, we show that in the vicinity of the points of concentration the density of optimal stationary distribution approximates the density of a Gaussian, and we explicitly solve for its covariance matrix.

CONTENTS

Received April 2015; revised October 2017.

[1]Supported in part by the Office of Naval Research through Grants N00014-14-1-0196 and N00014-16-1-2956, and in part by the Army Research Office through Grant W911NF-17-1-001.

[2]Supported in part by an award from the Simons Foundation (# 197982) to The University of Texas at Austin, in part by the Office of Naval Research through Grant N00014-14-1-0196, and in part by an INSPIRE faculty fellowship.

[3]Supported in part by a J. C. Bose Fellowship from the Department of Science and Technology, Government of India.

MSC2010 subject classifications. Primary 35R60; secondary 93E20.

Key words and phrases. Controlled diffusion, equilibrium selection, large deviations, small noise, ergodic control, HJB equation.

1. Introduction. The study of dynamical systems has a long and profound history. A lot of effort has been devoted to understand the behavior of the system when it is perturbed by an additive noise Berglund and Gentz (2006), Freidlin and Wentzell (1998), Olivieri and Vares (2005). Small noise diffusions have found applications in climate modeling Benzi et al. (1983), Berglund and Gentz (2002), electrical engineering Bobrovsky, Zakai and Zeitouni (1988), Zeitouni and Zakai (1992), finance Feng, Forde and Fouque (2010) and many other areas. Recent work on "stochastic resonance" [see, e.g., Moss (1994)] introduces an additional external input to the dynamics that may be viewed as a control. This is the main motivation for the study of the model we introduce next.

1.1. *The model.* In this paper, we consider a controlled dynamical system with small noise, which is modelled as a d-dimensional controlled diffusion $X = [X_1, \dots, X_d]^{\mathsf{T}}$ governed by the stochastic integral equation

$$X_t = X_0 + \int_0^t \bigl(m(X_s) + \varepsilon U_s\bigr)\,\mathrm{d}s + \varepsilon^{\nu} W_t, \qquad t \geq 0. \tag{1.1}$$

Here, all processes live in a complete probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ and the data of (1.1) satisfies the following.

(a) $m = [m_1, \dots, m_d]^{\mathsf{T}} \colon \mathbb{R}^d \to \mathbb{R}^d$ is a bounded $\mathcal{C}^{\infty}$ function with bounded derivatives.

(b) W is a standard Brownian motion in $\mathbb{R}^d$.

(c) U is an $\mathbb{R}^d$-valued control process which is jointly measurable in $(t, \omega) \in [0, \infty) \times \bar{\Omega}$ (in particular it has measurable paths), and is *nonanticipative*: for $t > s$, $W_t - W_s$ is independent of

$$\mathfrak{F}_s := \text{the completion of } \bigcap_{y > s} \sigma(X_0, W_r, U_r : r \leq y) \text{ relative to } (\mathfrak{F}, \mathbb{P}).$$

Such a control is called *admissible*, and we denote the set of admissible controls by $\mathfrak{U}$. As pointed out in Borkar (1989), p. 18, we may, without loss of generality, assume that an admissible U is adapted to the natural filtration of X.

(d) $0 < \varepsilon \ll 1$.

(e) $\nu > 0$.

Let $\mathcal{R}\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ be a *running cost* of the form

$$\mathcal{R}(x,u) := \ell(x) + \frac{1}{2}|u|^2, \tag{1.2}$$

where $\ell\colon \mathbb{R}^d \to \mathbb{R}_+$ is a prescribed smooth, Lipschitz function satisfying the condition

$$\lim_{|x|\to\infty} \ell(x) = \infty.$$

The control objective is to minimize the long run average (or *ergodic*) cost

$$\mathcal{J}^\varepsilon(U) := \limsup_{T\to\infty} \frac{1}{T}\,\mathbb{E}\Big[\int_0^T \mathcal{R}(X_s, U_s)\,\mathrm{d}s\Big], \tag{1.3}$$

over all admissible controls.

We view (1.1) as a perturbation of the o.d.e. (for *ordinary differential equation*)

$$\dot{x}(t) = m\big(x(t)\big), \tag{1.4}$$

perturbed by the "small noise" $\varepsilon^\nu W_t$ ("small" because $\varepsilon \ll 1$), and a control term εU_t. Since ε is small, the optimization criterion in (1.3) implies that the control is "expensive". We assume that the set of nonwandering points of the flow of (1.4) consists of finitely many hyperbolic equilibria, and that these are contained in some bounded open set which is positively invariant under the flow (see Hypothesis 1.1).

For the case when the control $U \equiv 0$, Freidlin and Wentzell (1998) developed a general framework for the analysis of small noise perturbed dynamical systems that is based on the theory of large deviations. Under a stochastic Lyapunov condition which we introduce later (Hypothesis 1.1), the cost is finite for $U = 0$, ensuring in particular that the set of controls $U \in \mathfrak{U}$ resulting in a finite value for $\mathcal{J}^\varepsilon(U)$ is nonempty. It is quite evident from ergodic theory that for $U = 0$ the limit (1.3) is the expectation of ℓ with respect to the invariant probability measure of (1.1).

The qualitative properties of the dynamics are best understood if we consider the special case $d = 1$, and $m = -\frac{\mathrm{d}F}{\mathrm{d}x}$ for some smooth function $F\colon \mathbb{R} \to \mathbb{R}$. Then the trajectory of (1.4) converges to a critical point of F. In fact, generically [i.e., for $x(0)$ in an open dense set] it converges to a stable one, that is, to a local minimum. If one views the graph of F as a "landscape", the local minima are the bottoms of its "valleys". The behavior of the stochastically perturbed (albeit uncontrolled)

version of this model, notably the study of the concentration of the stationary distribution, has been of considerable interest to physicists [see, e.g., Schuss (1980), Chapter 8 or Freidlin and Wentzell (1998), Chapter 6]. To find the actual support of the limit in the case of multiple equilibria, one often looks at the large deviation properties of these invariant measures Freidlin and Wentzell (1998). There are several studies in literature that deal with the large deviation principle of invariant measures of dynamical systems. Among the most relevant to the present are Sheu (1986) and Day (1987) which obtain a large deviation principle for invariant measures (more precisely, invariant densities) of (1.1) under the assumption that there is a unique equilibrium point. This has been extended to multiple equilibria in Biswas and Borkar (2009). A large deviation principle for invariant measures for a class of reaction-diffusion systems is established in Cerrai and Röckner (2005). However, none of the above mentioned studies have any control component in their dynamics.

The model in (1.1) goes a step further and considers the full-fledged optimal control version of this, wherein one tries to induce a preferred equilibrium behavior through a feedback control. The reason the latter has to be "expensive" is because this captures the physically realistic situation that one can "tweak" the dynamics but cannot replace it by something altogether different without incurring considerable expense. The function ℓ captures the relative preference among different points in the state space. Thus, the model in (1.1) is closely related to the model of stochastic resonance which has applications in neuron modelling, physics, electronics, physiology, etc. We refer to Herrmann et al. (2014), Chapter 1, for various applications in the presence of small noise. In particular, our model is closely related to the celebrated FitzHugh–Nagumo model Lindner, Bennett and Wiesenfeld (2006) in the presence of noise. The control in (1.1) should be seen as an external input. In practice it is convenient to take U to be periodic in time, whereas we do not impose any periodicity constraint on U. The ε factor in the control could be interpreted as the *weak modulation* in Moss (1994). We refer the reader to Moss (1994), Russell, Wilkens and Moss (1999) for a discussion on the interplay between noise variance and the control magnitude and its relation to stochastic resonance. Nonlinear control theory has been useful in understanding classes of systems that exhibit stochastic resonance Repperger and Farris (2010). Optimization theory has also been applied with the aim of enhancing the stochastic resonance effect for engineered systems Wu et al. (2006), Yang et al. (2009).

In our controlled setting, we are interested in achieving a desired value of β^{ε}_{*}, reflecting the desired behavior of the corresponding stationary distribution. Although one can fix a suitable penalty function ℓ beforehand, we will see in Theorem 1.11 in Section 1.4 that the value of β^{ε}_{*}, as well as the concentration of the stationary distribution, change with ν. Therefore, a desired value of β^{ε}_{*} or a desired profile of the stationary distribution might be obtained for some specific values of ν for small ε.

We also wish to point out that, since the control and noise are scaled differently, the ergodic control problem described can be viewed as a multi-scale diffusion problem.

1.1.1. *Assumptions on the vector field m.* Recall that a continuous-time dynamical system on a topological space $\mathcal{X}$ is specified by a map $\phi_t : \mathcal{X} \to \mathcal{X}$, where $\{\phi_t\}$ is a one parameter continuous Abelian group action on $\mathcal{X}$ called the *flow*. A point $x \in \mathcal{X}$ is called *nonwandering* if for every open neighborhood U of x and every time $T > 0$ there exists $t > T$ such that $\phi_t(U) \cap U \neq \varnothing$.

Recall also that a critical point z of a smooth vector field m is called *hyperbolic* if the Jacobian matrix $Dm(z)$ of m at z has no eigenvalues on the imaginary axis. For a hyperbolic critical point z of a vector field m, we let $\mathcal{W}_{\mathrm{s}}(z)$ and $\mathcal{W}_{\mathrm{u}}(z)$ denote the stable and unstable manifolds of its flow.

The following hypothesis on the vector field m is in effect throughout the paper.

HYPOTHESIS 1.1. The vector field m is bounded and smooth and satisfies the following:

1. The set of nonwandering points of the flow of m is a finite set $\mathcal{S} = \{z_1, \dots, z_n\}$ of hyperbolic critical points.
2. If y and z are critical points of m, then $\mathcal{W}_{\mathrm{s}}(y)$ and $\mathcal{W}_{\mathrm{u}}(z)$ intersect transversally (if they intersect).
3. There exist a smooth function $\bar{\mathcal{V}} \colon \mathbb{R}^d \to \mathbb{R}_+$ and a bounded open neighborhood of the origin $\mathcal{K} \subset \mathbb{R}^d$ containing $\mathcal{S}$, with the following properties.
   (3a) $c_1|x|^2 \le \bar{\mathcal{V}}(x) \le c_2(1 + |x|^2)$ for some positive constants c_1, c_2, and all $x \in \mathcal{K}^c$.
   (3b) $\nabla\bar{\mathcal{V}}$ is Lipschitz and satisfies

$$\langle m(x), \nabla\bar{\mathcal{V}}(x)\rangle < -\gamma|x| \tag{1.5}$$

   for some $\gamma > 0$, and all $x \in \mathcal{K}^c$.

REMARK 1.2. The vector field m is assumed bounded for simplicity. The reader however might notice that the characterization of optimality (see Theorem 1.4) is based on the regularity results in Bensoussan and Frehse (2002), and the hypotheses in Bensoussan and Frehse (2002), Section 4.6.1, permit m to be unbounded as long as

$$\limsup_{|x|\to\infty} \frac{|m(x)|^2}{\ell(x)} < \infty.$$

Provided that this condition is satisfied, the assumption that the drift is bounded can be waived and all the results of this paper hold unaltered, with the proofs requiring no major modification.

The outline of the paper is as follows. Section 1.2 summarizes the notation, and provides a glossary of special symbols used in the paper. In Section 1.5, we present an important property of LQG systems, which plays a crucial role in the study of the critical regime and also in the proof of Theorem 1.13.

In Section 2, we discuss energy functions for gradient-like flows (Theorem 2.2). These are heavily used in the study of the subcritical regime. The proofs of the main results comprise Sections 3–5. Section 3 is devoted to the study of the minimal stochastically stable sets, Section 4 is primarily devoted to the proof of Theorem 1.12, while Section 5 studies the optimal stationary distribution under an appropriate scaling, which leads to Theorem 1.13. The proofs of Lemma 1.3, Theorem 1.4, Lemma 1.17, and Theorem 1.19 can be found in Arapostathis, Biswas and Borkar (2018).

1.2. *Notation*. The following notation is used in this paper. The symbol $\mathbb{R}$ denotes the field of real numbers, and $\mathbb{N}$ denotes the set of natural numbers. The Euclidean norm on $\mathbb{R}^d$ is denoted by $|\cdot|$, and $\langle\cdot,\cdot\rangle$ denotes the inner product. For two real numbers a and b, $a \wedge b := \min(a,b)$ and $a \vee b := \max(a,b)$. For a matrix M, M^{T} denotes its transpose, and $\|M\|$ denotes the operator norm relative to the Euclidean vector norm. Also, I denotes the identity matrix.

The composition of two functions f and g is denoted by $f \circ g$. A ball of radius $r > 0$ in $\mathbb{R}^d$ around a point x is denoted by $B_r(x)$, or as B_r if $x = 0$. For a compact set K, we let $\operatorname{dist}(x,K)$ denote the Euclidean distance between $x \in \mathbb{R}^d$ and the set K, and $B_r(K) := \{y \in \mathbb{R}^d : \operatorname{dist}(y,K) < r\}$. For a set $A \subset \mathbb{R}^d$, we use $\bar{A}$, A^c, and ∂A to denote the closure, the complement, and the boundary of A, respectively. We define $\mathcal{C}_b^k(\mathbb{R}^d)$, $k \geq 0$, as the set of functions whose ith derivatives for $i = 0, 1, \dots, k$, are continuous and bounded in $\mathbb{R}^d$ and denote by $\mathcal{C}_c^k(\mathbb{R}^d)$ the subset of $\mathcal{C}_b^k(\mathbb{R}^d)$ of functions having compact support. The space of all probability measures on a Polish space $\mathcal{X}$ with the Prohorov topology is denoted by $\mathcal{P}(\mathcal{X})$. The density of the d-dimensional Gaussian distribution with mean 0 and covariance matrix Σ is denoted by ρ_Σ.

The term *domain* in $\mathbb{R}^d$ refers to a nonempty, connected open subset of the Euclidean space $\mathbb{R}^d$. We introduce the following notation for spaces of real-valued functions on a domain $G \subset \mathbb{R}^d$. The space $L^p(G)$, $p \in [1,\infty)$, stands for the usual Banach space of (equivalence classes of) measurable functions f satisfying $\int_G |f(x)|^p \,\mathrm{d}x < \infty$, and $L^\infty(G)$ is the Banach space of functions that are essentially bounded in G. The standard Sobolev space of functions on G whose generalized derivatives up to order k are in $L^p(G)$, equipped with its natural norm, is denoted by $\mathcal{W}^{k,p}(G)$, $k \geq 0$, $p \geq 1$.

In general, if $\mathcal{Y}$ is a space of real-valued functions on a domain G, $\mathcal{Y}_{\mathrm{loc}}$ consists of all functions f such that $f\varphi \in \mathcal{Y}$ for every $\varphi \in \mathcal{C}_c^\infty(G)$, the space of smooth functions on G with compact support. In this manner, we obtain for example the space $\mathcal{W}^{2,p}_{\mathrm{loc}}(G)$.

The symbols $\mathcal{O}(|x|^a)$ and $\mathfrak{o}(|x|^a)$, for $a \in (0,\infty)$, denote the sets of functions $f:\mathbb{R}^d \to \mathbb{R}$ having the property

$$\limsup_{|x|\searrow 0} \frac{|f(x)|}{|x|^a} < \infty \quad \text{and} \quad \limsup_{|x|\searrow 0} \frac{|f(x)|}{|x|^a} = 0,$$

respectively. Abusing the notation, $\mathcal{O}(|x|^a)$ and $\mathfrak{o}(|x|^a)$ occasionally denote generic members of these sets. Thus, for example, an inequality of the form $\mathcal{O}(|x|^2) \le f(x) \le \mathcal{O}(|x|)$ is well defined, and is equivalent to the statement that $\limsup_{|x|\searrow 0} \frac{|f(x)|}{|x|} < \infty$, and $\liminf_{|x|\searrow 0} \frac{f(x)}{|x|^2} > -\infty$.

We use $\kappa_1, \kappa_2, \dots$ as generic constants whose definition differs from place to place.

A glossary of commonly used symbols and the page where they are first defined is provided below.

Glossary of Symbols.

1.3. *The optimal stationary distribution.* Recall the function $\bar{\mathcal{V}}$ defined in Hypothesis 1.1. Since $\nabla\bar{\mathcal{V}}$ is Lipschitz, $\Delta\bar{\mathcal{V}}$ is bounded and thus (1.5) implies that with

$$\mathcal{L}_0^\varepsilon f(x) := \frac{\varepsilon^{2\nu}}{2}\Delta f(x) + \langle m(x), \nabla f(x)\rangle \qquad \forall x \in \mathbb{R}^d,\ f \in \mathcal{C}^2(\mathbb{R}^d),$$

we have

$$\mathcal{L}_0^\varepsilon \bar{\mathcal{V}}(x) \le \gamma_0 - \gamma|x| \qquad \forall \varepsilon \in (0,1),$$

for some positive constants γ and γ_0. This Foster–Lyapunov condition implies in particular that the process X with $U = 0$ has a unique invariant probability measure η_0^ε, and

$$\lim_{T\to\infty} \frac{1}{T}\,\mathbb{E}\biggl[\int_0^T |X_t|\,\mathrm{d}t\biggr] = \int_{\mathbb{R}^d} |x|\eta_0^\varepsilon(\mathrm{d}x) \le \frac{\gamma_0}{\gamma} \qquad \forall \varepsilon \in (0,1). \tag{1.6}$$

Since ℓ is Lipschitz, (1.6) implies that there exists a constant $\bar{c}_\ell$ independent of ε such that

$$\int \ell\,\mathrm{d}\eta_0^\varepsilon \le \bar{c}_\ell. \tag{1.7}$$

Moreover, from Biswas and Borkar (2009) there exists a unique Lipschitz continuous function $Z \ge 0$, such that $\min_{\mathbb{R}^d} Z = 0$, $Z(x) \to \infty$ as $|x| \to \infty$, and

$$Z(x) = \inf_{\phi:\,\phi(t)\to x_i, x_i\in\mathcal{S}} \biggl[\frac{1}{2}\int_0^\infty |\dot\phi(s) + m(\phi(s))|^2\,\mathrm{d}s + Z(x_i)\biggr], \qquad \phi(0) = x.$$

In addition, if ϱ_0^ε denotes the density of η_0^ε, then $-\varepsilon^{2\nu}\ln\varrho_0^\varepsilon(x) \to Z(x)$ uniformly on compact subsets of $\mathbb{R}^d$ as $\varepsilon \searrow 0$. The function Z is generally referred to as the *quasi-potential*, and plays a key role in the study of η_0^ε.

However, for the model in (1.1) under the optimal control criterion in (1.3), the standard method of analysis using quasi-potentials no longer applies. The first important step is to characterize the stationary probability distributions of the controlled diffusion under optimal controls. It is evident that optimal controls belong to the class $\widehat{\mathfrak{U}}$ defined by

$$\widehat{\mathfrak{U}} := \biggl\{U \in \mathfrak{U}\colon\ \mathbb{E}\biggl[\int_0^t |U_s|^2\,\mathrm{d}s\biggr] < \infty \text{ for all } t \ge 0\biggr\}. \tag{1.8}$$

We state the following result concerning the existence of solutions to (1.1). Its proof as well as the proof of Theorem 1.4 which appears later in this section can be found in the Supplementary Material [Arapostathis, Biswas and Borkar (2018)].

LEMMA 1.3. *Under any $U \in \widehat{\mathfrak{U}}$, the diffusion in (1.1) has a unique strong solution.*

1.3.1. *The convex analytic approach.* In studying this problem, it is of course of paramount importance to assert the existence of an optimal stationary distribution, and ideally, to also prove that it is unique.

A proper framework for this study is to consider the class $\mathfrak{P}^\varepsilon$ of *infinitesimal ergodic occupation measures*, that is, measures $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ which satisfy

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \mathcal{L}^\varepsilon[f](x,u)\pi(\mathrm{d}x,\mathrm{d}u) = 0 \qquad \forall f \in \mathcal{C}_{\mathrm{c}}^\infty(\mathbb{R}^d), \tag{1.9}$$

where $\mathcal{C}_{\mathrm{c}}^\infty(\mathbb{R}^d)$, as defined in Section 1.2, denotes the class of real-valued smooth functions with compact support. The operator $\mathcal{L}^\varepsilon : \mathcal{C}^2(\mathbb{R}^d) \to \mathcal{C}(\mathbb{R}^d \times \mathbb{R}^d)$ in (1.9) is defined by

$$\mathcal{L}^\varepsilon[f](x,u) := \frac{\varepsilon^{2\nu}}{2}\Delta f(x) + \bigl\langle m(x) + \varepsilon u, \nabla f(x)\bigr\rangle \tag{1.10}$$

for $f \in \mathcal{C}^2(\mathbb{R}^d)$. We adopt the usual relaxed control framework, where an admissible control is realized as a $\mathcal{P}(\mathbb{R}^d)$-valued measurable function [for details see Arapostathis, Borkar and Ghosh (2012), Section 2.3]. Thus if we disintegrate $\pi \in \mathfrak{P}^\varepsilon$ as $\pi(\mathrm{d}x,\mathrm{d}u) = \eta(\mathrm{d}x)\upsilon(\mathrm{d}u|x)$, and denote this as $\pi = \eta \circledast \upsilon$, then υ is a relaxed Markov control, and $\eta \in \mathcal{P}(\mathbb{R}^d)$ is an invariant probability measure for the corresponding controlled process, provided that the diffusion under the control υ in (1.1) has a unique weak solution for all $t \in [0,\infty)$ which is a Feller process.

Define

$$\mathcal{J}_\pi^\varepsilon := \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathcal{R}(x,u)\pi(\mathrm{d}x,\mathrm{d}u), \qquad \pi \in \mathfrak{P}^\varepsilon.$$

For a control $U \in \mathfrak{U}$ under which the diffusion has a unique weak solution we define the collection of *mean empirical measures* $\Phi_t^U \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ by

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} f(x,u)\Phi_t^U(\mathrm{d}x,\mathrm{d}u) = \frac{1}{t}\,\mathbb{E}\biggl[\int_0^t f(X_s,U_s)\,\mathrm{d}s\biggr] \tag{1.11}$$

for all $f \in \mathcal{C}_b(\mathbb{R}^d \times \mathbb{U})$.

Recall that a continuous function $f : \mathbb{R}^m \to \mathbb{R}$ is called *inf-compact* if the set $\{x \in \mathbb{R}^m : f(x) \le C\}$ is compact (or empty) for every $C \in \mathbb{R}$. Suppose that the ergodic cost $\mathcal{J}^\varepsilon(U)$ defined in (1.3) is finite. Then the inf-compactness of $\mathcal{R}(x,u)$ implies that the collection $\{\Phi_t^U, t > 0\}$ is tight in $\mathcal{P}(\mathbb{R}^d \times \mathbb{U})$. It is standard to show, by following an argument similar to the proof of Lemma 3.4.6 in Arapostathis, Borkar and Ghosh (2012), that any limit point $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{U})$ of Φ_t^U is an infinitesimal ergodic occupation measure. Moreover, $\mathcal{J}^\varepsilon(U) \ge \inf_{\pi \in \mathfrak{P}^\varepsilon} \mathcal{J}_\pi^\varepsilon$ [Arapostathis, Borkar and Ghosh (2012), Theorem 3.4.6]. It is natural then to consider the convex minimization problem

$$\mathcal{J}_*^\varepsilon := \inf_{\pi \in \mathfrak{P}^\varepsilon} \mathcal{J}_\pi^\varepsilon, \tag{1.12}$$

since $\mathcal{J}^\varepsilon_*$ provides a lower bound for $\mathcal{J}^\varepsilon(U)$. This constitutes the *primal problem*. Since $\mathcal{R}(x,u)$ is inf-compact, $\pi \mapsto \mathcal{J}^\varepsilon(\pi)$ is lower semi-continuous, and $\mathcal{J}^\varepsilon_\pi$ is finite for at least one $\pi \in \mathfrak{P}^\varepsilon$ by (1.7), it follows that there exists some $\pi^\varepsilon_* \in \mathfrak{P}^\varepsilon$ which attains the infimum in (1.12). If the disintegration of an optimal ergodic occupation measure results in a Markov control under which (1.1) has a solution, then of course this infimum is attained for the ergodic control problem. This is indeed the case, for a large class of problems where the control takes values in a compact space. For general results concerning this approach see Bhatt and Borkar (1996) and Kurtz and Stockbridge (1998). However, for problems when the control lives in $\mathbb{R}^d$, as is the case in the present setup, it is in general difficult to show that under the Markov control associated with π^ε_* the diffusion has a solution.

The dual of the infinite dimensional linear program in (1.12) consists of a maximization over subsolutions of a HJB equation Bhatt and Borkar (1996). We say that we have *strong duality* if the optimal values of the primal and the dual problems are equal. To the best of our knowledge, there are no strong duality results for the ergodic control problem of diffusions when the control lives in $\mathbb{R}^d$. In the next section, we study the HJB equation and establish strong duality for the problem at hand. Moreover, we establish the unicity of the optimal ergodic occupation measure $\pi^\varepsilon_* = \eta^\varepsilon_* \circledast v^\varepsilon_*$. This of course implies that there exist a unique "optimal" stationary distribution η^ε_* and an a.e. unique optimal stationary Markov control, and it turns out from the study of the HJB that this control is smooth.

1.3.2. *The HJB equation for the ergodic control problem.* Recall that a *precise* stationary Markov control is specified as $U_t = v(X_t)$ for a measurable function $v\colon \mathbb{R}^d \to \mathbb{R}^d$. We identify the stationary Markov control with the function v. Let $\mathfrak{U}_{\mathrm{SM}}$ denote the class of stationary Markov controls which are locally bounded and under which (1.1) has a unique strong solution for all $t \in [0,\infty)$. Parenthetically, we note that, under a locally bounded stationary Markov control, (1.1) has a unique solution up to explosion time, and it is strong Feller [Krylov and Röckner (2005), Theorem 2.5]. Linear growth of $|v|$ is a sufficient condition for the existence of a unique strong solution for all $t \in [0,\infty)$. We let $\mathbb{E}^v_x$ denote the expectation operator on the canonical space of the process controlled by $v \in \mathfrak{U}_{\mathrm{SM}}$, and starting at $X_0 = x$. We say that $v \in \mathfrak{U}_{\mathrm{SM}}$ is *stable* if the controlled process under v is positive recurrent, and we let $\mathfrak{U}^\varepsilon_{\mathrm{SSM}} \subset \mathfrak{U}_{\mathrm{SM}}$ denote the set of *stable* controls in $\mathfrak{U}_{\mathrm{SM}}$. Parts (a)–(b) of the following theorem essentially follow from Ichihara (2011), Theorem 2.2.

THEOREM 1.4. *There exists a critical value $\beta^\varepsilon_* \in \mathbb{R}$ such that the HJB equation for the ergodic control problem, given by*

$$\frac{\varepsilon^{2\nu}}{2}\Delta V^\varepsilon + \min_{u\in\mathbb{R}^d}\Big[\big\langle m + \varepsilon u, \nabla V^\varepsilon\big\rangle + \ell + \frac{1}{2}|u|^2\Big] = \beta^\varepsilon, \tag{1.13}$$

has no solution if $\beta^\varepsilon > \beta^\varepsilon_*$, while if $\beta^\varepsilon < \beta^\varepsilon_*$, then for any such solution V^ε the diffusion in (1.1) under the control $v = -\varepsilon\nabla V^\varepsilon$ is transient. Moreover, the following hold.

(a) *If $V^\varepsilon \in \mathcal{C}^2(\mathbb{R}^d)$ is any solution of (1.13), then $|\nabla V^\varepsilon(x)|$ has at most affine growth in x.*

(b) *If $\beta^\varepsilon = \beta^\varepsilon_*$, then (1.13) has a unique solution $V^\varepsilon \in \mathcal{C}^2(\mathbb{R}^d)$ satisfying $V^\varepsilon(0) = 0$. The Markov control $v^\varepsilon_* := -\varepsilon\nabla V^\varepsilon$ is stable, and if $\eta^\varepsilon_* \in \mathcal{P}(\mathbb{R}^d)$ denotes the invariant probability measure of the diffusion under the control v^ε_*, then*

$$\beta^\varepsilon_* = \int_{\mathbb{R}^d} \mathcal{R}\bigl(x, v^\varepsilon_*(x)\bigr)\eta^\varepsilon_*(\mathrm{d}x). \tag{1.14}$$

(c) (*strong duality*) $\mathcal{J}^\varepsilon_* = \beta^\varepsilon_*$.

(d) *With $\widehat{\mathfrak{U}}$ as defined in (1.8), the following optimality property holds*:

$$\liminf_{T\to\infty} \inf_{U\in\widehat{\mathfrak{U}}} \frac{1}{T}\,\mathbb{E}\biggl[\int_0^T \mathcal{R}(X_s, U_s)\,\mathrm{d}s\biggr] \ge \beta^\varepsilon_*.$$

(e) (*uniqueness of optimal stationary distribution*) *An ergodic occupation measure $\pi = \eta \circledast v \in \mathfrak{P}^\varepsilon$ is optimal if and only if v agrees with v^ε_* a.e. in $\mathbb{R}^d$. In particular, there exists a unique optimal invariant probability measure η^ε_*.*

For a stationary Markov control v, we define the *extended generator* of (1.1) by

$$\mathcal{L}^\varepsilon_v f(x) := \frac{\varepsilon^{2\nu}}{2}\Delta f(x) + \bigl\langle m(x) + \varepsilon v(x), \nabla f(x)\bigr\rangle, \qquad x \in \mathbb{R}^d, \tag{1.15}$$

for $f \in \mathcal{C}^2(\mathbb{R}^d)$. It follows from (1.13) that

$$\frac{\varepsilon^{2\nu}}{2}\Delta V^\varepsilon + \bigl\langle m, \nabla V^\varepsilon\bigr\rangle - \frac{\varepsilon^2}{2}\bigl|\nabla V^\varepsilon\bigr|^2 + \ell = \beta^\varepsilon_*. \tag{1.16}$$

Theorem 1.4 shows that $\beta^\varepsilon_* = \mathcal{J}^\varepsilon_*$, and this value is attained at an a.e. unique $v^\varepsilon_* \in \mathfrak{U}^\varepsilon_{\mathrm{SSM}}$ and is independent of the initial condition X_0. Given these uniqueness properties, we refer to η^ε_* as *the optimal invariant probability measure*, or as *the optimal stationary distribution*, and we let ϱ^ε_* denote its density. We also refer to v^ε_* as *the optimal stationary Markov control*, and to β^ε_* as *the optimal value* for the ergodic problem.

REMARK 1.5. Due to the smoothness of coefficients, every weak solution in $V^\varepsilon \in \mathcal{W}^{1,\infty}_{\mathrm{loc}}(\mathbb{R}^d)$ of (1.13) is automatically in $\mathcal{C}^k(\mathbb{R}^d)$ for any $k \in \mathbb{N}$. In the interest of notational economy, we often refer to any such V^ε as a solution, without specifying the function space it belongs to.

REMARK 1.6. Existence and uniqueness of the solution to (1.13) is well known [see Bensoussan and Frehse (1992, 2002)], and in fact, the results in

Bensoussan and Frehse (2002) hold for a more general class of HJB equations. However, we were not able to find any reference that establishes the verification of optimality results in Theorem 1.4, or strong duality.

Note also that Theorem 1.4(d) asserts a much stronger optimality property than the usual one. This can be in fact strengthened to pathwise optimality, and assert that the most "pessimistic" pathwise performance under v_*^ε is no worse than the most "optimistic" pathwise performance under any control in $\widehat{\mathfrak{U}}$. The proof of this fact is identical to the proofs of Lemma 3.4.6 and Theorem 3.4.7 in Arapostathis, Borkar and Ghosh (2012).

Recent work as in Ichihara (2012) and Ichihara and Sheu (2013) which investigates the optimal control problem, does not exactly fit our model. A strict growth condition for ℓ is imposed in Assumption (H2) of Ichihara (2012), which we do not require here. On the other hand, in Ichihara and Sheu (2013) where convergence of the Cauchy problem is investigated, and therefore optimality for the ergodic control problem is addressed, a more stringent condition is imposed [see Hypothesis (A3)′] which, for a Hamiltonian that is quadratic in the gradient like ours, amounts to geometric ergodicity under the uncontrolled dynamics.

The existence of a critical value for β^ε for (1.13) and the behavior of the solutions above or below this critical value are studied in detail in Ichihara (2011). However, the critical value is not necessarily the optimal value. We refer the reader to Ichihara (2015) for some recent work on the relation of the critical value of an elliptic HJB equation of the ergodic type and the optimal value of the control problem.

1.4. *Main results.* In this section we summarize the main results of the paper. We start with the following definition.

DEFINITION 1.7. Let $\mathcal{S}_{\mathrm{s}} \subset \mathcal{S}$ denote the set of stable equilibria of (1.4), that is, the set of points $z \in \mathcal{S}$ for which the eigenvalues of $Dm(z)$ have negative real parts.

We say that a set $K \subset \mathbb{R}^d$ is *stochastically stable* (or that η_*^ε *concentrates* on K) if it is compact, and for any open neighborhood $\mathcal{N} \supset K$ we have $\lim_{\varepsilon \searrow 0} \eta_*^\varepsilon(\mathcal{N}) = 1$. If $\mathfrak{H}$ denotes the class of stochastically stable sets, and $\mathfrak{S} := \bigcap_{K \in \mathfrak{H}} K$, then $\mathfrak{S}$ is stochastically stable (Remark 1.8). We refer to $\mathfrak{S}$ as the *minimal stochastically stable set*.

REMARK 1.8. It is straightforward to show that $\mathfrak{S}$ in Definition 1.7 is stochastically stable. This goes as follows. For a set $K \subset \mathbb{R}^d$, and $\delta > 0$, let K^δ denote the open δ-neighborhood of K, that is, $K^\delta := \{x \in \mathbb{R}^d : d(x, K) < \delta\}$, where $d(\cdot, \cdot)$ is the Euclidean distance. Since the collection $\mathfrak{H}$ consists of compact sets, it follows there exists a finite subcollection $K_1^\delta, \dots, K_n^\delta$ whose intersection lies in $\mathfrak{S}^{2\delta}$. Then $\eta_*^\varepsilon((\mathfrak{S}^{2\delta})^c) \le \bigcup_{i=1}^n \eta_*^\varepsilon((K_i^\delta)^c)$, from which it follows, since $\delta > 0$ is arbitrary, that $\mathfrak{S}$ is stochastically stable.

The behavior of η^ε_* for small ε depends crucially on the parameter ν. We distinguish three regimes: The *supercritical regime* ($\nu > 1$), the *subcritical regime* ($\nu < 1$), and the *critical regime* ($\nu = 1$). Roughly speaking, the control "exceeds" the noise level in the supercritical regime, while the opposite is the case in the subcritical regime. In the critical regime, which is the most interesting and more difficult to study, the control and noise levels are equal. The main results can be grouped in three categories: (1) characterization of the minimal stochastically stable set $\mathfrak{S}$ and asymptotic estimates of β^ε_* for small ε in the three regimes (Theorem 1.11), (2) concentration bounds for η^ε_* (Theorem 1.12), and (3) convergence of ϱ^ε_*, under appropriate scaling, to a Gaussian density (Theorem 1.13).

DEFINITION 1.9. For a square matrix $M \in \mathbb{R}^{d\times d}$, let $\Lambda^+(M)$ denote the sum of its eigenvalues that lie in the open right half complex plane. For $z \in \mathcal{S}$, and with $M_z := Dm(z)$, where as defined earlier $Dm(z)$ is the Jacobian of m at z, we let $\widehat{Q}_z$ and $\widehat{\Sigma}_z$ be the symmetric, nonnegative definite, square matrices solving the pair of equations

$$\begin{aligned} M_z^{\mathsf{T}}\widehat{Q}_z + \widehat{Q}_z M_z &= \widehat{Q}_z^2, \\ (M_z - \widehat{Q}_z)\widehat{\Sigma}_z + \widehat{\Sigma}_z(M_z - \widehat{Q}_z)^{\mathsf{T}} &= -I. \end{aligned} \tag{1.17}$$

By Theorem 1.19, which appears in Section 1.5, there exists a unique pair $(\widehat{Q}_z, \widehat{\Sigma}_z)$ of symmetric positive semidefinite matrices solving (1.17). It is also evident by (1.17) that $\widehat{\Sigma}_z$ is invertible.

In order to state the main results, we need the following definition.

DEFINITION 1.10. We define the *optimal control effort* $\mathcal{G}^\varepsilon_*$ by

$$\mathcal{G}^\varepsilon_* := \frac{1}{2}\int_{\mathbb{R}^d} |v^\varepsilon_*|^2 \,\mathrm{d}\eta^\varepsilon_*, \qquad \varepsilon > 0. \tag{1.18}$$

In addition, we define

$$\begin{aligned} \mathcal{Z}_{\mathrm{c}} &:= \operatorname*{Arg\,min}_{z\in\mathcal{S}}\{\ell(z) + \Lambda^+(Dm(z))\}, \qquad \mathfrak{J}_{\mathrm{c}} := \min_{z\in\mathcal{S}}[\ell(z) + \Lambda^+(Dm(z))], \\ \mathcal{Z}_{\mathrm{s}} &:= \operatorname*{Arg\,min}_{z\in\mathcal{S}_{\mathrm{s}}}\{\ell(z)\}, \qquad \mathfrak{J}_{\mathrm{s}} := \min_{z\in\mathcal{S}_{\mathrm{s}}}[\ell(z)], \\ \mathcal{Z} &:= \operatorname*{Arg\,min}_{z\in\mathcal{S}}\{\ell(z)\}, \qquad \mathfrak{J} := \min_{z\in\mathcal{S}}[\ell(z)], \\ \widetilde{\mathcal{Z}} &:= \operatorname*{Arg\,min}_{z\in\mathcal{Z}}\{\Lambda^+(Dm(z))\}, \qquad \widetilde{\mathfrak{J}} := \min_{z\in\mathcal{Z}}[\Lambda^+(Dm(z))]. \end{aligned}$$

Recall the definition of $\mathcal{O}(\cdot)$ in Section 1.2. The following theorem provides a comprehensive characterization of the minimal stochastically stable set.

THEOREM 1.11. *The minimal stochastically stable set $\mathcal{S}$ is a subset of $\mathcal{S}$ for all $\nu > 0$. In addition, the set $\mathcal{S}$, the optimal value β^ε_*, and the optimal control effort $\mathcal{G}^\varepsilon_*$ depend on ν as follows.*

(i) *For $\nu > 1$ ("supercritical" regime), we have $\mathcal{S} \subset \widetilde{\mathcal{Z}}$. In addition, if $\mathfrak{J} = \mathfrak{J}_{\rm s}$, then*

$$\mathcal{O}\big(\varepsilon^{2\wedge\nu}\big) \le \beta^\varepsilon_* - \mathfrak{J} \le \mathcal{O}\big(\varepsilon^{2\nu}\big) \quad \text{and} \quad \mathcal{G}^\varepsilon_* \in \mathcal{O}\big(\varepsilon^{\nu\wedge 2}\big),$$

and if $\mathfrak{J} < \mathfrak{J}_{\rm s}$, then

$$\mathcal{O}\big(\varepsilon^{2\wedge\nu}\big) \le \beta^\varepsilon_* - \mathfrak{J} \le \varepsilon^{2\nu-2}\widetilde{\mathfrak{J}} + \mathcal{O}\big(\varepsilon^{2\nu}\big) \quad \text{and} \quad \mathcal{G}^\varepsilon_* \in \mathcal{O}\big(\varepsilon^{(2\nu-2)\wedge 2}\big).$$

(ii) *For $\nu < 1$ ("subcritical" regime), we have $\mathcal{S} \subset \mathcal{Z}_{\rm s}$, and*

$$\mathcal{O}\big(\varepsilon^{\nu}\big) \le \beta^\varepsilon_* - \mathfrak{J}_{\rm s} \le \mathcal{O}\big(\varepsilon^{\nu\vee(4\nu-2)}\big), \qquad \mathcal{G}^\varepsilon_* \in \mathcal{O}\big(\varepsilon^{\nu}\big). \tag{1.19}$$

(iii) *For $\nu = 1$ ("critical" regime), we have $\mathcal{S} \subset \mathcal{Z}_{\rm c}$, $\beta^\varepsilon_* \le \mathfrak{J}_{\rm c} + \mathcal{O}(\varepsilon^2)$, and $\lim_{\varepsilon\searrow 0} \beta^\varepsilon_* = \mathfrak{J}_{\rm c}$. Moreover, if $\mathfrak{J}_{\rm c} = \mathfrak{J}_{\rm s}$, then the lower bound in (1.19) holds.*

It is not hard to show that the optimal invariant measures η^ε_* concentrate on $\mathcal{S}$ as $\varepsilon \searrow 0$ (see Lemma 3.1). In Theorem 1.11, we distinguish the three regimes corresponding to different values of ν, and provide asymptotic bounds for β^ε_* for small ε. For $\nu > 1$, one can find a control U under which the invariant measure of the dynamics (1.1) concentrates on a point in $\mathcal{S}$. Construction of invariant measures with similar properties is also possible for $z \in \mathcal{S}_{\rm s}$ when $\nu < 1$. The important difference is that for $\nu < 1$ the optimal invariant measure η^ε_* cannot concentrate on $\mathcal{S} \setminus \mathcal{S}_{\rm s}$ (see Lemma 3.6). To show this fact we construct a suitable energy function for the Morse–Smale dynamics (see Theorem 2.2). The analysis in the critical regime $\nu = 1$ turns out to be more subtle than the other two regimes. To facilitate the study of the critical regime, we identify an important property which concerns a singular ergodic control problem for Linear Quadratic Gaussian (LQG) systems (Theorem 1.19). This plays a crucial role in showing that $\mathcal{S} \subset \mathcal{Z}_{\rm c}$.

To guide the reader, we indicate the results presented in Sections 3–4 which comprise the proof of Theorem 1.11.

PROOF OF THEOREM 1.11. That $\mathcal{S} \subset \mathcal{S}$ is the statement of Lemma 3.1. Note that if $\mathfrak{J} = \mathfrak{J}_{\rm s}$, then $\widetilde{\mathfrak{J}} = \min_{z\in\mathcal{Z}}[\Lambda^+(Dm(z))] = 0$ by the definition of Λ^+. Thus upper bounds of $\beta^\varepsilon_* - \mathfrak{J}$ in part (i) follow by the first inequality in (3.17), while the lower bounds are in Corollary 4.2(b). The statements concerning $\mathcal{G}^\varepsilon_*$ in part (i) are in (4.4).

That $\mathcal{S} \subset \mathcal{Z}_{\rm s}$ in the subcritical regime is in the statement of Lemma 3.6. The upper bound of $\beta^\varepsilon_* - \mathfrak{J}_{\rm s}$ in part (ii) is the combination of the two separate upper bounds given in Lemma 3.5(ii), for $\nu \in (0, \frac{2}{3})$ and $\nu \in [\frac{2}{3}, 1)$, while the lower bound is in Corollary 4.6(b), where we also find the assertion that $\mathcal{G}^\varepsilon_* \in \mathcal{O}(\varepsilon^\nu)$.

We now turn to the proof of part (iii). The inequality $\beta_*^\varepsilon \le \mathfrak{J}_c + \mathcal{O}(\varepsilon^2)$ is the second inequality in (3.17). That $\lim_{\varepsilon\searrow 0}\beta_*^\varepsilon = \mathfrak{J}_c$ is in the statement of Theorem 5.4, while $\mathfrak{S}\subset\mathcal{Z}_c$ is equivalent to $\lim_{\varepsilon\searrow 0}\eta_*^\varepsilon(B_r^c(\mathcal{Z}_c)) = 0$, which is asserted in (5.12). Lastly, that $\beta_*^\varepsilon - \mathfrak{J}_s \ge \mathcal{O}(\varepsilon)$ when $\mathfrak{J}_c = \mathfrak{J}_s$ is in Remark 4.7. □

The next theorem provides concentration bounds for the optimal stationary distribution in terms of moments. Let $\operatorname{dist}(x,\mathcal{S})$ denote the Euclidean distance of $x\in\mathbb{R}^d$ from the set $\mathcal{S}$, and $B_r(\mathcal{S}) := \{y\in\mathbb{R}^d\colon \operatorname{dist}(y,\mathcal{S}) < r\}$.

THEOREM 1.12. *For any $k\in\mathbb{N}$ and $r>0$, there exist constants, $\hat\kappa_0 = \hat\kappa_0(k,r,\nu)$, and $\hat\kappa_i = \hat\kappa_i(k)$, $i=1,2$, such that with $\hat r(\varepsilon) := \hat\kappa_2\varepsilon^{\nu\wedge 1}$ we have*

$$\begin{aligned}\int_{B_r(\mathcal{S})}\big(\operatorname{dist}(x,\mathcal{S})\big)^2\eta_*^\varepsilon(\mathrm{d}x) &\le \hat\kappa_0\varepsilon^{2(\nu\wedge 2)} \qquad \forall \nu>0,\\ \int_{B^c_{\hat r(\varepsilon)}(\mathcal{S})}\big(\operatorname{dist}(x,\mathcal{S})\big)^{2k}\eta_*^\varepsilon(\mathrm{d}x) &\le \hat\kappa_1\varepsilon^{2(\nu\wedge 1)} \qquad \forall \nu\in(0,2],\end{aligned} \tag{1.20}$$

for all $\varepsilon\in(0,1)$.

Moreover, if D is any open set such that $\mathcal{S}_s\subset D$, then

$$\eta_*^\varepsilon(D^c)\in\mathcal{O}\big(\varepsilon^{2\nu\wedge(2-\nu)}\big),$$

provided $\nu<1$, or $\mathfrak{J}_c=\mathfrak{J}_s$ and $\nu=1$, or $\mathfrak{J}=\mathfrak{J}_s$ and $\nu\in(1,2)$.

PROOF. The first inequality in (1.20) is the same as (4.1), while the second is established in Proposition 4.5.

That $\eta_*^\varepsilon(D^c)\in\mathcal{O}(\varepsilon^{2\nu\wedge(2-\nu)})$ when $\mathfrak{J}=\mathfrak{J}_s$ and $\nu\in(1,2)$, or when $\nu<1$ is asserted in Corollary 4.6, and that the same inclusion holds when $\mathfrak{J}_c=\mathfrak{J}_s$ and $\nu=1$ is explained in Remark 4.7. □

Exploiting the results in Theorem 1.12, we scale the space suitably and show that the resulting invariant measures are also tight. In particular, we examine the asymptotic behavior of η_*^ε and show that under an appropriate spatial scaling it "converges" to a Gaussian distribution in the vicinity of the minimal stochastically stable set. This is the subject of the next theorem.

THEOREM 1.13. *Assume $\nu\in(0,2)$. Let $z\in\mathcal{S}$, and $\mathcal{N}$ an open neighborhood of z whose closure does not contain any other elements of $\mathcal{S}$. Suppose that along some sequence $\varepsilon_n\searrow 0$ we have $\liminf_{\varepsilon_n\searrow 0}\eta_*^{\varepsilon_n}(\mathcal{N})>0$. Then along this sequence it holds that*

$$\frac{\varepsilon^{\nu d}\varrho_*^\varepsilon(\varepsilon^\nu x+z)}{\eta_*^\varepsilon(\mathcal{N})}\xrightarrow[\varepsilon\searrow 0]{}\frac{1}{(2\pi)^{\frac{d}{2}}|\det\widehat\Sigma_z|^{\frac12}}\exp\Big(-\frac12\langle x,\widehat\Sigma_z^{-1}x\rangle\Big), \tag{1.21}$$

*uniformly on compact sets, where "*det*" denotes the determinant, and $\widehat\Sigma_z$ is given by* (1.17).

PROOF. This follows from Theorems 5.3 and 5.7. □

We present a simple example to demonstrate the results.

EXAMPLE 1.14. Let m be a vector field in $\mathbb{R}$ of the form $m = -\nabla F$, with F a "double well potential" given by $F(x) := \frac{x^4}{4} - \frac{x^3}{3} - x^2$ on $[-10, 10]$, with F suitably extended so that it is globally Lipschitz and does not have any critical points outside the interval $[-10, 10]$. Then ∇F vanishes at exactly three points: $-1, 0, 2$. Of these, 0 is a local maximum, and therefore, it is an unstable equilibrium for the o.d.e. $\dot{x}(t) = m(x(t))$, and both -1 and 2 are local minima, hence stable equilibria thereof. Let $\ell(x) = c|x|^2$ on $[-10, 10]$ for a suitable $c > 0$, modified suitably outside $[-10, 10]$ to render it globally Lipschitz. Note that $F(0) = 0$, $F(-1) = -\frac{5}{12}$, $F(2) = -\frac{8}{3}$. Thus $x = 2$ is the unique global minimum of F. Since $\ell(0) = 0$, and $Dm(0) = 2$, the results of Theorem 1.11 indicate that for ε small enough the following hold:

- in the supercritical regime $\mathfrak{S} = \{0\}$, and $\beta^\varepsilon_* \approx \ell(0) = 0$;
- in the subcritical regime $\mathfrak{S} = \{-1\}$, $\beta^\varepsilon_* \approx \ell(-1) = c$;
- in the critical regime, we have $\mathfrak{S} = \{0\}$ if $c > 2$, with $\beta^\varepsilon_* \approx \ell(0) + Dm(0) = 2$, and $\mathfrak{S} = \{-1\}$ if $c < 2$, with $\beta^\varepsilon_* \approx \ell(-1) = c$.

Next, we change the data so that

$$F(x) := \frac{x^6}{6} - \frac{x^5}{5} - \frac{7x^4}{4} + \frac{x^3}{3} + 3x^2 \qquad \text{on } [-10, 10].$$

Then ∇F vanishes at exactly five points, and $\mathcal{S} = \{-2, -1, 0, 1, 3\}$. Of these, -1 and 1 are local maxima of F, hence unstable equilibria for the o.d.e. $\dot{x}(t) = m(x(t))$, while the rest are stable equilibria. Hence, $\mathcal{S}_{\mathrm{s}} = \{-2, 0, 3\}$. Let $\ell(x) = 5x^4 - x^3 - 20x^2 + 16$ on $[-10, 10]$. The critical point $z = 3$ is the unique global minimum for F, which means that it is stochastically stable for the uncontrolled dynamics. Calculating the values of ℓ at $\mathcal{S}$ we obtain $\ell(-2) = 24$, $\ell(-1) = 2$, $\ell(0) = 16$, $\ell(1) = 0$, and $\ell(3) = 214$. Also, we have $Dm(-1) = 8$, $Dm(1) = 12$. By Theorem 1.11, we have the following:

- in the supercritical regime, $\mathfrak{S} = \{1\}$, and $\beta^\varepsilon_* \approx \ell(1) = 0$ for ε small;
- in the critical regime, $\mathfrak{S} = \{-1\}$, and $\beta^\varepsilon_* \approx \ell(-1) + Dm(-1) = 10$ for ε small;
- in the subcritical regime, $\mathfrak{S} = \{0\}$, $\beta^\varepsilon_* \approx \ell(0) = 16$ for ε small.

Note that in this example the stochastically stable sets are distinct in the three regimes.

REMARK 1.15. Theorems 1.11–1.12 suggest that $\nu = 2$ is a critical value. We present an example with linear drift and quadratic penalty, so that explicit calculations are possible, to show that indeed $\nu = 2$ is a critical value. Consider a one-dimensional model with data $m(x) = x$ and $\ell(x) = (x + 1)^2$. Direct substitution

shows that the solution of the HJB equation [see (1.16)] is

$$V^{\varepsilon}(x) = \frac{1+\sqrt{1+2\varepsilon^2}}{2\varepsilon^2}\biggl(x + \frac{2\varepsilon^2}{(1+\sqrt{1+2\varepsilon^2})\sqrt{1+2\varepsilon^2}}\biggr)^2,$$

$$\beta_*^{\varepsilon} = \frac{1}{1+2\varepsilon^2} + \varepsilon^{2\nu-2}\frac{1+\sqrt{1+2\varepsilon^2}}{2}.$$

The closed loop drift is

$$\begin{aligned} x - \varepsilon^2 \nabla V^{\varepsilon}(x) &= -\sqrt{1+2\varepsilon^2}x - \frac{2\varepsilon^2}{\sqrt{1+2\varepsilon^2}} \\ &= -\sqrt{1+2\varepsilon^2}\biggl(x + \frac{2\varepsilon^2}{1+2\varepsilon^2}\biggr). \end{aligned} \tag{1.22}$$

Thus, the optimal stationary distribution η_*^{ε} is Gaussian with variance $(\sigma_*^{\varepsilon})^2$ and mean $\mathfrak{m}_*^{\varepsilon}$ given by

$$(\sigma_*^{\varepsilon})^2 := \frac{\varepsilon^{2\nu}}{2\sqrt{1+2\varepsilon^2}}, \qquad \mathfrak{m}_*^{\varepsilon} := -\frac{2\varepsilon^2}{1+2\varepsilon^2}. \tag{1.23}$$

Consider the scaled distribution $\hat{\eta}_*^{\varepsilon}$ with density $\varepsilon^{\nu}\varrho_*^{\varepsilon}(\varepsilon^{\nu}x+z)$. Let $\mathcal{N}(\mathfrak{m},\sigma^2)$ denote the Normal distribution with mean $\mathfrak{m}$ and variance σ^2. We have:

- For $\nu \in (0,2)$, $\hat{\eta}_*^{\varepsilon}$ converges to $\mathcal{N}(0,\frac{1}{2})$.
- For $\nu = 2$, $\hat{\eta}_*^{\varepsilon}$ converges to $\mathcal{N}(-2,\frac{1}{2})$.
- For $\nu > 2$, we have $\frac{\mathfrak{m}_*^{\varepsilon}}{\sigma_*^{\varepsilon}} \to -\infty$, and thus $\hat{\eta}_*^{\varepsilon}$ does not converge as $\varepsilon \searrow 0$.

Thus (1.21) does not hold for $\nu \geq 2$.

A simple calculation also shows that the optimal control effort is given by

$$\begin{aligned} \mathcal{G}_*^{\varepsilon} &= \frac{\varepsilon^{-2}}{2}\bigl(1+\sqrt{1+2\varepsilon^2}\bigr)^2(\sigma_*^{\varepsilon})^2 \\ &\quad + \frac{\varepsilon^{-2}}{2}\bigl(1+\sqrt{1+2\varepsilon^2}\bigr)^2\biggl(\frac{2\varepsilon^2}{(1+\sqrt{1+2\varepsilon^2})\sqrt{1+2\varepsilon^2}} + \mathfrak{m}_*^{\varepsilon}\biggr)^2 \\ &= \varepsilon^{2\nu-2}\frac{(1+\sqrt{1+2\varepsilon^2})^2}{4\sqrt{1+2\varepsilon^2}} + \frac{2\varepsilon^2}{(1+2\varepsilon^2)^2}. \end{aligned}$$

Thus $\mathcal{G}_*^{\varepsilon} \in \mathcal{O}(\varepsilon^{(2\nu-2)\wedge 2})$, which matches the estimate in Theorem 1.11(i).

A better understanding of this can be reached by considering the limit $\nu \to \infty$, in which case the dynamics are deterministic. A simple calculation shows that

$$\bar{x} := \arg\min_x \biggl\{\ell(\varepsilon x) + \frac{1}{2}|x|^2\biggr\} = -\frac{2\varepsilon^2}{1+2\varepsilon^2}.$$

Thus for a feedback control to be optimal, the point $\bar{x}$ should be asymptotically stable for the closed loop system. As a result, for the LQG problem, the optimal stationary distribution is centered at the point $\bar{x}$ for all values of ν. The criticality at $\nu = 2$ is generic, since in the vicinity of an equilibrium z, solving the minimization problem we have $\bar{x} \approx \varepsilon^2 \nabla \ell(z)$.

There is a similar behavior if the drift is stable. Let $m(x) = -x$. We obtain

$$V^\varepsilon(x) = \frac{-1+\sqrt{1+2\varepsilon^2}}{2\varepsilon^2}\biggl(x + \frac{2\varepsilon^2}{(-1+\sqrt{1+2\varepsilon^2})\sqrt{1+2\varepsilon^2}}\biggr)^2,$$

$$\begin{aligned}\beta^\varepsilon_* &= \frac{1}{1+2\varepsilon^2} + \varepsilon^{2\nu-2}\frac{-1+\sqrt{1+2\varepsilon^2}}{2} \\ &= 1 - \frac{2\varepsilon^2}{1+2\varepsilon^2} + \varepsilon^{2\nu}\frac{1}{1+\sqrt{1+2\varepsilon^2}}.\end{aligned}$$

The closed loop drift, variance, and mean are as in (1.22)–(1.23). Using the identity

$$\frac{-1+\sqrt{1+2\varepsilon^2}}{2\varepsilon^2} = \frac{1}{1+\sqrt{1+2\varepsilon^2}},$$

the optimal control effort takes the form

$$\begin{aligned}\mathcal{G}^\varepsilon_* &= \frac{2(\sigma^\varepsilon_*)^2}{(1+\sqrt{1+2\varepsilon^2})^2} + \frac{2\varepsilon^2}{(1+\sqrt{1+2\varepsilon^2})^2}\biggl(\frac{1+\sqrt{1+2\varepsilon^2}}{\sqrt{1+2\varepsilon^2}} + \mathfrak{m}^\varepsilon_*\biggr)^2 \\ &= \frac{\varepsilon^{2\nu}}{(1+\sqrt{1+2\varepsilon^2})^2\sqrt{1+2\varepsilon^2}} + \frac{2\varepsilon^2}{(1+2\varepsilon^2)^2}.\end{aligned}$$

Thus $\mathcal{G}^\varepsilon_* \in \mathcal{O}(\varepsilon^{2\nu\wedge 2})$.

1.5. *A property of LQG systems.* As mentioned earlier, the study of the critical regime, and also the proof of Theorem 1.13 rely on an important property of LQG systems which we describe next. A matrix $M \in \mathbb{R}^{d\times d}$ is called *exponentially dichotomous* if it has no eigenvalues on the imaginary axis. Consider the diffusion

$$\mathrm{d}X_t = \bigl(MX_t + v(X_t)\bigr)\,\mathrm{d}t + \mathrm{d}W_t, \tag{1.24}$$

with $M \in \mathbb{R}^{d\times d}$ exponentially dichotomous.

DEFINITION 1.16. Let $\overline{\mathfrak{U}}_{\mathrm{SSM}}$ denote the class of locally bounded stationary Markov controls v, under which the diffusion in (1.24) has a unique strong solution, is positive recurrent, and satisfies

$$\mathcal{E}(v) := \frac{1}{2}\int_{\mathbb{R}^d} \bigl|v(x)\bigr|^2 \mu_v(\mathrm{d}x) < \infty, \tag{1.25}$$

where μ_v denotes the associated invariant probability measure.

As Theorem 1.19 below asserts, the minimal control effort, defined by

$$\mathcal{E}_* := \inf_{v \in \overline{\mathfrak{U}}_{\mathrm{SSM}}} \mathcal{E}(v),$$

which is required to render the diffusion positive recurrent by controls in $\overline{\mathfrak{U}}_{\mathrm{SSM}}$, equals the trace of the unstable spectrum of the matrix M, which was denoted as $\Lambda^+(M)$ in Definition 1.9. This result is related to classical results in deterministic linear control systems and the Riccati equation Kučera (1972), Mårtensson (1971), Willems (1971), but since we could not locate it in this form in the literature, a proof is included in the Supplementary Material [Arapostathis, Biswas and Borkar (2018)], where the proof of the following auxiliary lemma is also located.

LEMMA 1.17. *Provided M is exponentially dichotomous, there exists a constant $\widetilde{C}_0$ depending only on M such that*

$$\int_{\mathbb{R}^d} |x|^2 \mu_v(\mathrm{d}x) \le \widetilde{C}_0 \Big(1 + \int_{\mathbb{R}^d} |v(x)|^2 \mu_v(\mathrm{d}x) \Big) \qquad \forall v \in \overline{\mathfrak{U}}_{\mathrm{SSM}}.$$

Recall that a real square matrix is called *Hurwitz* if its eigenvalues lie in the open left half complex plane. We need the following definition.

DEFINITION 1.18. Let $M \in \mathbb{R}^{d\times d}$ be fixed. Let $\mathcal{G}(M)$ denote the collection of all matrices $G \in \mathbb{R}^{d\times d}$ such that $M - G$ is Hurwitz. For $G \in \mathcal{G}(M)$, let Σ_G denote the (unique) symmetric solution of the Lyapunov equation

$$(M - G)\Sigma_G + \Sigma_G (M - G)^{\mathsf{T}} = -I, \tag{1.26}$$

and define

$$\begin{aligned} \mathcal{J}_G(M) &:= \frac{1}{2} \operatorname{trace}\big(G \Sigma_G G^{\mathsf{T}}\big), \\ \mathcal{J}_*(M) &:= \inf_{G \in \mathcal{G}(M)} \mathcal{J}_G(M). \end{aligned} \tag{1.27}$$

Let $v_G(x) = -Gx$ for some $G \in \mathbb{R}^{d\times d}$. It is clear that for the diffusion in (1.24) to be positive recurrent under the linear control v_G, it is necessary that $M - G$ be Hurwitz. If so, then the invariant probability distribution of the controlled diffusion is Gaussian with covariance matrix Σ_G given by (1.26). It is clear then that the control effort $\mathcal{E}(v_G)$ defined in (1.25) satisfies $\mathcal{E}(v_G) = \mathcal{J}_G(M)$. Therefore, provided the infimum in (1.27) is attained, $\mathcal{J}_*(M)$ is the minimal control effort, as defined by (1.25), required to render (1.24) positive recurrent using a linear stationary Markov control. Theorem 1.19 asserts that the infimum in (1.27) is indeed attained and that $\mathcal{J}_*(M) = \Lambda^+(M)$. Moreover, linear stationary Markov controls are optimal for this task within the class $\overline{\mathfrak{U}}_{\mathrm{SSM}}$.

THEOREM 1.19. *Suppose that $M \in \mathbb{R}^{d\times d}$ is exponentially dichotomous. Then the following hold.*

(a) *There exists a unique positive semidefinite symmetric solution Q of the matrix Riccati equation $M^{\mathsf{T}}Q + QM = Q^2$, satisfying*

$$(M-Q)\Sigma + \Sigma(M-Q)^{\mathsf{T}} = -I \tag{1.28}$$

for some symmetric positive definite matrix Σ. Moreover, $A = M - Q$ attains the infimum in (1.27) subject to (1.26), and it holds that

$$\mathcal{J}_*(M) = \Lambda^+(M) = \frac{1}{2}\operatorname{trace}(Q).$$

(b) *With μ_v denoting the invariant probability measure of (1.24) under a control $v \in \overline{\mathfrak{U}}_{\mathrm{SSM}}$, we have*

$$\inf_{v\in\overline{\mathfrak{U}}_{\mathrm{SSM}}} \int_{\mathbb{R}^d} \frac{1}{2}\bigl|v(x)\bigr|^2 \mu_v(\mathrm{d}x) = \Lambda^+(M). \tag{1.29}$$

In addition, any control $v_* \in \overline{\mathfrak{U}}_{\mathrm{SSM}}$ which attains the infimum in (1.29) satisfies $v_*(x) = -Qx$ for almost all x in $\mathbb{R}^d$.

(c) *Let $\bar\beta \in \mathbb{R}$. The equation*

$$\frac{1}{2}\Delta\bar V(x) + \langle Mx, \nabla\bar V(x)\rangle - \frac{|\nabla\bar V(x)|^2}{2} = \bar\beta \tag{1.30}$$

has no solution if $\bar\beta > \Lambda^+(M)$. If $\bar\beta = \Lambda^+(M)$, then $\bar V(x) = \frac{1}{2}\langle x, Qx\rangle$ is the unique solution of (1.30) satisfying $\bar V(0) = 0$. If $\bar\beta < \Lambda^+(M)$ and $\bar V$ is a solution of (1.30), then the diffusion in (1.24) under the control $v = -\nabla\bar V$ is transient.

REMARK 1.20. Optimality and uniqueness of the optimal control $v(x) = -Qx$ in Theorem 1.19(b) hold over a larger class of Markov controls. Indeed combining the results of Bogachev, Röckner and Shaposhnikov (2012), Krylov and Röckner (2005), we can replace "locally bounded" in the definition of $\overline{\mathfrak{U}}_{\mathrm{SSM}}$ by $v \in L^p_{\mathrm{loc}}(\mathbb{R}^d)$ for some $p > d$. Then the results of Theorem 1.19(b) hold for this class of controls.

2. Gradient-like flows and energy functions.

2.1. *Gradient-like Morse–Smale dynamical systems.* It is well known in the theory of dynamical systems that if the set of nonwandering points of a flow on a compact manifold consists of hyperbolic fixed points, then the associated vector field is generically *gradient-like* (see Definition 2.1 and Theorem 2.2 below). This is also the case under Hypothesis 1.1, since the "point at infinity" is a source for the flow of m.

Recall that the *index* of a hyperbolic critical point $z \in \mathbb{R}^d$ of a smooth vector field is defined as the dimension of the unstable manifold $\mathcal{W}_u(z)$. This agrees with the number of eigenvalues of $Dm(z)$ which have positive real parts. The theorem below is well known Meyer (1968), Smale (1961). What we have added in its statement is the assertion that the energy function can be chosen in a manner that its Laplacian at critical points of the vector field with positive index is negative.

We start with the following definition.

DEFINITION 2.1. We say that $\mathcal{V} \in \mathcal{C}^\infty(\mathbb{R}^d)$ is an *energy function* if it is inf-compact, and has a finite set $\mathcal{S} = \{z_1, \dots, z_n\}$ of critical points, which are all non-degenerate. A $\mathcal{C}^\infty$ vector field m on $\mathbb{R}^d$ is called *gradient-like relative to* an energy function $\mathcal{V}$ provided that every point in $\mathcal{S}$ is a hyperbolic critical point of m, and

$$\langle m(x), \nabla \mathcal{V}(x) \rangle < 0 \qquad \forall x \in \mathbb{R}^d \setminus \mathcal{S}.$$

If m satisfies these properties, we also say that m is *adapted to* $\mathcal{V}$.

THEOREM 2.2. *Suppose that m is a smooth vector field in $\mathbb{R}^d$ for which Hypothesis 1.1 holds. Let G be any domain of $\mathbb{R}^d$ of the form $\{x \in \mathbb{R}^d : \bar{\mathcal{V}} < c\}$ for some $c \in \mathbb{R}$, satisfying $G \supset \mathcal{K}$, and let $\{a_z : z \in \mathcal{S}\}$ be any set of distinct real numbers such that if z and z' are the α- and ω-limit points of some trajectory, respectively, then $a_z > a_{z'}$. Then there exists a function $\widehat{\mathcal{V}} \in \mathcal{C}^\infty(\bar{G})$ with the following properties.*

(i) $\langle m(x), \nabla \widehat{\mathcal{V}}(x) \rangle < 0$ *for all* $x \in \bar{G} \setminus \mathcal{S}$.

(ii) *For each $z \in \mathcal{S}$, there exist a neighborhood $\mathcal{N}_z$ of z and a symmetric matrix $Q_z \in \mathbb{R}^{d \times d}$ such that $\widehat{\mathcal{V}}(x) = a_z + \langle x - z, Q_z(x - z) \rangle + \mathfrak{o}(|x - z|^2)$ for all $x \in \mathcal{N}_z$.*

(iii) $\Delta \widehat{\mathcal{V}}(z) < 0$, *for all $z \in \mathcal{S} \setminus \mathcal{S}_s$, where $\mathcal{S}_s$, as defined earlier, denotes the stable equilibria of the flow of m.*

(iv) *There exists a constant $C_0 > 0$ such that*

$$\begin{aligned} C_0 \big(\operatorname{dist}(x, \mathcal{S}) \vee |\nabla \widehat{\mathcal{V}}(x)|\big)^2 &\le |\langle m(x), \nabla \widehat{\mathcal{V}}(x) \rangle| \\ &\le C_0^{-1} \big(\operatorname{dist}(x, \mathcal{S}) \wedge |\nabla \widehat{\mathcal{V}}(x)|\big)^2 \end{aligned} \tag{2.1}$$

for all $x \in G$.

PROOF. Since m is smooth and bounded, and $m(z) = 0$ for $z \in \mathcal{S}$, there exists a constant $\tilde{C}_m > 0$ such that

$$|M_z x - m(x)| \le \tilde{C}_m |x|^2 \qquad \forall x \in \mathbb{R}^d, \forall z \in \mathcal{S}. \tag{2.2}$$

Let $z \in \mathcal{S}$ be a critical point of m of index $q \ge 0$. Translating the coordinates we may assume that $z = 0$. Since $m(0) = 0$, then by (2.2), $m(x)$ takes the form

$$m(x) = Mx + \mathcal{O}\big(|x|^2\big)$$

locally around $x = 0$, where $M = Dm(0)$. By hypothesis M has exactly q $(d - q)$ eigenvalues in the open right half (left half) complex space. Therefore, since the corresponding eigenspaces are invariant under M, there exists a linear coordinate transformation T such that, in the new coordinates $\tilde{x} = T(x)$, the linear map $x \mapsto Mx$ has the matrix representation $\tilde{M} = TMT^{-1}$ and $\tilde{M} = \operatorname{diag}(\tilde{M}_1, -\tilde{M}_2)$, where $\tilde{M}_1$ and $\tilde{M}_2$ are square Hurwitz matrices of dimension $d-q$ and q respectively. By the Lyapunov theorem, there exist positive definite matrices $\tilde{Q}_i$, $i = 1, 2$, satisfying

$$\begin{aligned} \tilde{M}_1^{\mathsf{T}} \tilde{Q}_1 + \tilde{Q}_1 \tilde{M}_1 &= -I_{d-q}, \\ \tilde{M}_2^{\mathsf{T}} \tilde{Q}_2 + \tilde{Q}_2 \tilde{M}_2 &= -I_q, \end{aligned} \tag{2.3}$$

where I_{d-q} and I_q are the identity matrices of dimension $d-q$ and q, respectively. Suppose $q > 0$, and let $\theta > 1$ be such that

$$\theta \operatorname{trace}\bigl(T^{\mathsf{T}} \operatorname{diag}(0, \tilde{Q}_2) T\bigr) > \operatorname{trace}\bigl(T^{\mathsf{T}} \operatorname{diag}(\tilde{Q}_1, 0) T\bigr), \tag{2.4}$$

and define $\widehat{\mathcal{V}}$ in some neighborhood of 0 by

$$\widehat{\mathcal{V}}(x) := a + \bigl\langle x, T^{\mathsf{T}} \operatorname{diag}(\tilde{Q}_1, -\theta \tilde{Q}_2) T x \bigr\rangle, \tag{2.5}$$

where a is a constant to be determined later. By (2.4), we obtain $\Delta \widehat{\mathcal{V}}(0) < 0$, and thus (iii) holds.

Using (2.2), we have

$$\begin{aligned} \bigl\langle m(x), \nabla \widehat{\mathcal{V}}(x) \bigr\rangle = x^{\mathsf{T}} \bigl[& M^{\mathsf{T}} T^{\mathsf{T}} \operatorname{diag}(\tilde{Q}_1, -\theta \tilde{Q}_2) T \\ & + T^{\mathsf{T}} \operatorname{diag}(\tilde{Q}_1, -\theta \tilde{Q}_2) T M \bigr] x + \mathcal{O}\bigl(|x|^3\bigr). \end{aligned}$$

Expanding this, we obtain

$$\begin{aligned} T^{\mathsf{T}} \operatorname{diag}(\tilde{Q}_1, -\theta \tilde{Q}_2) T M &= T^{\mathsf{T}} \operatorname{diag}(\tilde{Q}_1, -\theta \tilde{Q}_2) T T^{-1} \tilde{M} T \\ &= T^{\mathsf{T}} \operatorname{diag}(\tilde{Q}_1 \tilde{M}_1, \theta \tilde{Q}_2 \tilde{M}_2) T. \end{aligned}$$

By (2.3), we obtain

$$\bigl\langle m(x), \nabla \widehat{\mathcal{V}}(x) \bigr\rangle = -\bigl\langle x, T^{\mathsf{T}} \operatorname{diag}(I_{d-q}, \theta I_q) T x \bigr\rangle + \mathcal{O}\bigl(|x|^3\bigr).$$

Therefore, since $\theta > 1$, we have

$$-\theta |Tx|^2 + \mathcal{O}\bigl(|x|^3\bigr) \le \bigl\langle m(x), \nabla \widehat{\mathcal{V}}(x) \bigr\rangle \le -|Tx|^2 + \mathcal{O}\bigl(|x|^3\bigr). \tag{2.6}$$

As shown in Smale (1961) one can select any real numbers a_i and define $\widehat{\mathcal{V}}$ on $\mathcal{S}$ by setting $\widehat{\mathcal{V}}(z_i) = a_i$ as long as the following consistency condition is met. If z_i and z_j are the α- and ω-limit points of some trajectory, then $a_i > a_j$. Thus $\widehat{\mathcal{V}}$ can be defined in nonoverlapping neighborhoods of the critical points by (2.5) so as to satisfy (2.6) and parts (i)–(iii) of the theorem. Since G is positively invariant under the flow of m, the stable and unstable manifolds of $\mathcal{S}$ intersect transversally by Hypothesis 1.1(2), and m is transversal to the boundary of ∂G by Hypothesis 1.1(3b),

this function can then be extended to $\bar{G}$ by the handlebody decomposition technique introduced by Smale. For details see Smale (1961), Theorem B, and Meyer (1968), Theorem 1.

It is clear by (2.5)–(2.6) that (2.1) holds in some open neighborhood of each $z \in \mathcal{S}$, and thus, $\mathcal{S}$ being a finite set, it also holds in some neighborhood $\mathcal{N}$ of $\mathcal{S}$. Since $\langle m, \nabla\widehat{\mathcal{V}}\rangle$ is strictly negative on the compact set $\bar{G} \setminus \mathcal{N}$ and $\langle m(x), \nabla\widehat{\mathcal{V}}(x)\rangle < 0$ for all $x \notin \mathcal{S}$, a constant C_0 can be selected so that (2.1) holds on G. This completes the proof. □

The function $\widehat{\mathcal{V}}$ in Theorem 2.2 can be extended to $\mathbb{R}^d$, and constructed in a manner so that it agrees, outside some ball, with the Lyapunov function $\bar{\mathcal{V}}$ in Hypothesis 1.1. This is stated in the following lemma.

LEMMA 2.3. *Under the assumptions of Theorem 2.2 the vector field m is adapted to an energy function $\mathcal{V}$ which satisfies $\mathcal{V} = \bar{\mathcal{V}}$ on the complement of some open ball which contains $\mathcal{S}$. In addition, parts* (i)–(iv) *of Theorem 2.2 hold, and for every bounded domain G there exists a constant $C_0 = C_0(G)$ such that* (2.1) *holds for all $x \in G$. Moreover, there exists a constant $\overline{C}_0 > 0$ such that with*

$$\overline{\mathcal{V}}(x) := \max\bigl\{\bigl(\operatorname{dist}(x,\mathcal{S})\bigr)^2 \wedge \operatorname{dist}(x,\mathcal{S}), \bigl|\nabla\mathcal{V}(x)\bigr|^2 \wedge \bigl|\nabla\mathcal{V}(x)\bigr|\bigr\},$$

$$\underline{\mathcal{V}}(x) := \min\bigl\{\bigl(\operatorname{dist}(x,\mathcal{S})\bigr)^2 \wedge \operatorname{dist}(x,\mathcal{S}), \bigl|\nabla\mathcal{V}(x)\bigr|^2 \wedge \bigl|\nabla\mathcal{V}(x)\bigr|\bigr\},$$

we have

$$\overline{C}_0^{-1}\overline{\mathcal{V}}(x) \le \bigl|\langle m(x), \nabla\mathcal{V}(x)\rangle\bigr| \le \overline{C}_0\underline{\mathcal{V}}(x) \qquad \forall x \in \mathbb{R}^d. \tag{2.7}$$

PROOF. Select $c \in \mathbb{R}$ such that $G_1 := \{x \in \mathbb{R}^d : \bar{\mathcal{V}} < c\}$ is a domain which contains $\mathcal{K}$. Let $G_2 := \{x \in \mathbb{R}^d : \bar{\mathcal{V}} < 2c\}$. By Theorem 2.2 there exists $\widehat{\mathcal{V}} \in \mathcal{C}^\infty(G_2)$ with the properties stated. Without loss of generality, we can assume that $\widehat{\mathcal{V}} = 2c$ on ∂G_2 [Smale (1961), Theorem B]. Let $c_1 := \sup_{G_1}\widehat{\mathcal{V}}$. It follows that $c_1 < 2c$ by the positive invariance of G_2, and the property $\langle m, \nabla\widehat{\mathcal{V}}\rangle < 0$ in $G_2 \setminus G_1$. We write $A \Subset B$ to indicate that $\bar{A} \subset B$. Let

$$\tilde{G} := \left\{x \in \mathbb{R}^d : \widehat{\mathcal{V}} < \frac{(c_1 + 2c)}{2}\right\},$$

and $c_2 := \sup_{\tilde{G}}\bar{\mathcal{V}}$. Then we have $G_1 \Subset \tilde{G} \Subset G_2$, and $c < c_2 < 2c$ by construction.

Let $\psi : \mathbb{R} \to \mathbb{R}$ be a smooth nondecreasing function such that $\psi(t) = t$ for $t \le \frac{1}{2}(c_1 + 2c)$, $\psi(t) = 2c$ for $t \ge 2c$, and whose derivative is strictly positive on the interval $[\frac{1}{2}(c_1 + 2c), 2c)$. Similarly, let $\bar{\psi} : \mathbb{R} \to \mathbb{R}$ be another smooth nondecreasing function such that $\bar{\psi}(t) = 0$ for $t \le -c$, and $\bar{\psi}(t) = t$ for $t \ge c_2 - 2c$. Define

$$\mathcal{V} := \psi \circ \widehat{\mathcal{V}} + \bar{\psi} \circ (\bar{\mathcal{V}} - 2c).$$

By construction, $\mathcal{V}$ agrees with $\widehat{\mathcal{V}}$ on G_1 and with $\bar{\mathcal{V}}$ on G_2^c. It can also be easily verified that $\sup_{G_2\setminus G_1}\langle m, \nabla\mathcal{V}\rangle < 0$. Thus, $\mathcal{V}\in\mathcal{C}^\infty(\mathbb{R}^d)$ is an energy function, and m is adapted to $\mathcal{V}$ according to Definition 2.1.

Since $\langle m(x), \nabla\mathcal{V}(x)\rangle < 0$ for all $x\notin\mathcal{S}$, and $\mathcal{V}$ agrees with $\widehat{\mathcal{V}}$ on $\mathcal{K}$, parts (i)–(iv) of Theorem 2.2 clearly hold. Also, since (2.7) holds in some neighborhood of $\mathcal{S}$ by (2.5)–(2.6), then, in view of the linear growth of $\langle m(x), \nabla\bar{\mathcal{V}}(x)\rangle \neq 0$ in (1.5), and the assumptions on the growth of $\bar{\mathcal{V}}$ in Hypothesis 1.1, the inequalities in (2.7) have to hold on $\mathbb{R}^d$. □

3. Minimal stochastically stable sets. Recall that β_*^ε denotes the optimal value of (1.3), η_*^ε denotes the stationary distribution of the process X under the optimal stationary Markov control v_*^ε, and ϱ_*^ε denotes its density. These definitions are fixed throughout the rest of the paper. Recall also the definition of the extended generator in (1.15), and the definition of $\mathcal{R}$ in (1.2). For a stationary Markov control v, we use the notation

$$\mathcal{R}[v](x) := \mathcal{R}\big(x, v(x)\big) = \ell(x) + \frac{1}{2}\big|v(x)\big|^2. \tag{3.1}$$

Throughout the rest of the paper, $\mathcal{V}$ is a smooth function which satisfies (i)–(iv) in Theorem 2.2 and agrees with $\bar{\mathcal{V}}$ in Hypothesis 1.1 on the complement of some open ball which contains $\mathcal{S}$ (Lemma 2.3). We refer to $\mathcal{V}$ as the *energy function*.

We start the analysis with the following lemma which asserts that η_*^ε concentrates on $\mathcal{S}$ as $\varepsilon\searrow 0$.

LEMMA 3.1. *The family* $\{\eta_*^\varepsilon, \varepsilon\in(0,1)\}$ *is tight, and any sub-sequential limit as* $\varepsilon\searrow 0$ *has support on* $\mathcal{S}$.

PROOF. Recall that η_0^ε denotes the invariant probability measure of (1.1) under the control $U=0$. Define

$$\beta_0^\varepsilon := \int_{\mathbb{R}^d} \ell(x)\eta_0^\varepsilon(\mathrm{d}x).$$

By (1.7) we have

$$\int_{\mathbb{R}^d} \ell(x)\eta_*^\varepsilon(\mathrm{d}x) \le \beta_*^\varepsilon \le \beta_0^\varepsilon \le \bar{c}_\ell \qquad \forall\varepsilon\in(0,1). \tag{3.2}$$

Since ℓ is inf-compact, (3.2) implies that $\{\eta_*^\varepsilon, \varepsilon\in(0,1)\}$ is tight. Let $\phi_t(x)$ denote the solution of (1.4) starting at $x\in\mathbb{R}^d$ at $t=0$, that is, $\phi_0(x)=x$. If C_m denotes a Lipschitz constant of m and $X_0=x$, we have

$$\big|X_t-\phi_t(x)\big| \le C_m\int_0^t \big|X_s-\phi_s(x)\big|\,\mathrm{d}s + \varepsilon\int_0^t \big|v_*^\varepsilon(X_s)\big|\,\mathrm{d}s + \varepsilon^\nu|W_t|. \tag{3.3}$$

Hence, applying Gronwall's inequality, we obtain from (3.3) that

$$\sup_{s\in[0,t]}\big|X_s-\phi_s(x)\big| \le \mathrm{e}^{C_m t}\left(\varepsilon\int_0^t \big|v_*^\varepsilon(X_s)\big|\,\mathrm{d}s + \varepsilon^\nu \sup_{s\le t}|W_s|\right). \tag{3.4}$$

In turn, for any $\delta > 0$, (3.4) implies that

$$\mathbb{P}_x\big(|X_t - \phi_t(x)| \geq \delta\big) \leq \mathbb{P}_x\bigg(\int_0^t |v_*^\varepsilon(X_s)|\,\mathrm{d}s \geq \frac{\delta \mathrm{e}^{-C_m t}}{2\varepsilon}\bigg) + \mathbb{P}_x\bigg(\sup_{s\leq t}|W_s| \geq \frac{\delta \mathrm{e}^{-C_m t}}{2\varepsilon^\nu}\bigg)$$

for $t > 0$. By Jensen's inequality we obtain

$$\begin{aligned}\mathbb{P}_x\bigg(\int_0^t |v_*^\varepsilon(X_s)|\,\mathrm{d}s \geq \frac{\delta \mathrm{e}^{-C_m t}}{2\varepsilon}\bigg) &\leq \mathbb{P}_x\bigg(\int_0^t |v_*^\varepsilon(X_s)|^2\,\mathrm{d}s \geq \frac{\delta^2 \mathrm{e}^{-2C_m t}}{4t\varepsilon^2}\bigg)\\ &\leq \frac{4t\varepsilon^2}{\delta^2}\mathrm{e}^{2C_m t}\,\mathbb{E}_x\bigg[\int_0^t |v_*^\varepsilon(X_s)|^2\,\mathrm{d}s\bigg].\end{aligned}$$

Therefore, for any compact set $K \subset \mathbb{R}^d$ we have

$$\begin{aligned}\int_K \mathbb{P}_x\big(|X_t - \phi_t(x)| \geq \delta\big)\eta_*^\varepsilon(\mathrm{d}x) &\leq \frac{4t^2\varepsilon^2}{\delta^2}\mathrm{e}^{2C_m t}\int_{\mathbb{R}^d}|v_*^\varepsilon(x)|^2\eta_*^\varepsilon(\mathrm{d}x)\\ &\quad + \sup_{x\in K}\mathbb{P}_x\bigg(\sup_{s\leq t}|W_s| \geq \frac{\delta}{2\varepsilon^\nu}\mathrm{e}^{-C_m t}\bigg).\end{aligned} \tag{3.5}$$

It is clear that the right-hand side of (3.5) tends to 0 as $\varepsilon \searrow 0$. Thus, for any compact set $K \subset \mathbb{R}^d$, and any Lipschitz function $f \in \mathcal{C}_b(\mathbb{R}^d)$ it holds that

$$\int_K \big|\mathbb{E}_x^{v_*^\varepsilon}\big[f(X_t)\big] - f\big(\phi_t(x)\big)\big|\eta_*^\varepsilon(\mathrm{d}x) \xrightarrow[\varepsilon\searrow 0]{} 0. \tag{3.6}$$

On the other hand, since η_*^ε is an invariant probability measure, we have

$$\int_{\mathbb{R}^d}\mathbb{E}_x^{v_*^\varepsilon}\big[f(X_t)\big]\eta_*^\varepsilon(\mathrm{d}x) = \int_{\mathbb{R}^d} f(x)\eta_*^\varepsilon(\mathrm{d}x) \qquad \forall f \in \mathcal{C}_b\big(\mathbb{R}^d\big), \forall t \geq 0. \tag{3.7}$$

Let $\bar\eta \in \mathcal{P}(\mathbb{R}^d)$ be any limit of η_*^ε along some sequence $\{\varepsilon_n\}$, with $\varepsilon_n \searrow 0$ as $n \to \infty$. By (3.6)–(3.7), the tightness of $\{\eta_*^\varepsilon, \varepsilon \in (0,1)\}$, and a standard triangle inequality, we obtain

$$\int_{\mathbb{R}^d} f\big(\phi_t(x)\big)\bar\eta(\mathrm{d}x) = \int_{\mathbb{R}^d} f(x)\bar\eta(\mathrm{d}x) \qquad \forall t \geq 0, \tag{3.8}$$

for all Lipschitz functions $f \in \mathcal{C}_b(\mathbb{R}^d)$. Since the ω-limit set of any trajectory of (1.4) is contained in $\mathcal{S}$, equation (3.8) shows that $\bar\eta$ has support on $\mathcal{S}$. This completes the proof. □

3.1. *Two lemmas concerning the case $\nu \geq 1$.* For $z \in \mathcal{S}$, let $\bar v_z^\varepsilon$ for $\varepsilon \in (0,1)$, denote the stationary Markov control defined by

$$\bar v_z^\varepsilon(x) := \frac{(M_z - \widehat{Q}_z)(x - z) - m(x)}{\varepsilon}, \qquad t \geq 0, \tag{3.9}$$

where M_z and $\widehat{Q}_z$ are as in Definition 1.9. The controlled process is then governed by the diffusion

$$\mathrm{d}X_t = (M_z - \widehat{Q}_z)(X_t - z)\,\mathrm{d}t + \varepsilon^{\nu}\,\mathrm{d}W_t. \tag{3.10}$$

Since $M_z - \widehat{Q}_z$ is Hurwitz by Theorem 1.19, the diffusion has a stationary probability distribution $\bar{\mu}^{\varepsilon}_z$, which is Gaussian with mean z and covariance matrix $\varepsilon^{2\nu}\widehat{\Sigma}_z$, with $\widehat{\Sigma}_z$ given in (1.17).

We start with the following lemma.

LEMMA 3.2. *Suppose that $\nu \ge 1$ and $z \in \mathcal{S}$. Let $\bar{v}^{\varepsilon}_z$ be the stationary Markov control in* (3.9), *and $\bar{\mu}^{\varepsilon}_z$ the invariant probability measure of the diffusion governed by* (3.10). *Then*

$$\begin{aligned} \int_{\mathbb{R}^d} \frac{1}{2}\bigl|\bar{v}^{\varepsilon}(x)\bigr|^2 \bar{\mu}^{\varepsilon}_z(\mathrm{d}x) &= \varepsilon^{2\nu-2}\Lambda^+\bigl(Dm(z)\bigr) + \mathcal{O}\bigl(\varepsilon^{4\nu-2}\bigr), \\ \int_{\mathbb{R}^d} \ell(x)\bar{\mu}^{\varepsilon}_z(\mathrm{d}x) &= \ell(z) + \mathcal{O}\bigl(\varepsilon^{2\nu}\bigr). \end{aligned} \tag{3.11}$$

PROOF. Without loss of generality, we assume that $z = 0$, and simplifying the notation, we let $M = M_z$, $Q = \widehat{Q}_z$, $\Sigma = \widehat{\Sigma}_z$, and $\bar{\mu}^{\varepsilon} = \bar{\mu}^{\varepsilon}_z$. Then we have

$$\bigl|(M-Q)x - m(x)\bigr|^2 = |Qx|^2 + 2\bigl\langle Qx, Mx - m(x)\bigr\rangle + \bigl|Mx - m(x)\bigr|^2. \tag{3.12}$$

Since by Taylor's theorem it holds that

$$\bigl\langle Qx, Mx - m(x)\bigr\rangle = \bigl\langle Qx, F(x)\bigr\rangle + \mathcal{O}\bigl(|x|^4\bigr),$$

with

$$F(x) := \bigl(F_1(x), \dots, F_d(x)\bigr) \quad \text{and} \quad F_i(x) := \frac{1}{2}\bigl\langle x, \nabla^2 m_i(0)x\bigr\rangle,$$

using (2.2) and (3.12) we obtain

$$\bigl|(M-Q)x - m(x)\bigr|^2 = |Qx|^2 + 2\bigl\langle Qx, F(x)\bigr\rangle + \mathcal{O}\bigl(|x|^4\bigr). \tag{3.13}$$

As mentioned in the paragraph preceding the lemma, $\bar{\mu}^{\varepsilon}$ is Gaussian, with zero mean, and covariance matrix $\varepsilon^{2\nu}\Sigma$, where Σ is the solution of (1.28). Since $\langle Qx, F(x)\rangle$ is a homogeneous polynomial of degree 3, it has zero mean under the Gaussian. It is also the case that the fourth moments of $\bar{\mu}^{\varepsilon}$ are of order $\varepsilon^{4\nu}$. It then follows by the estimate in (3.13) and Theorem 1.19(b) that

$$\begin{aligned} \frac{1}{2}\int_{\mathbb{R}^d} \bigl|\bar{v}^{\varepsilon}(x)\bigr|^2 \bar{\mu}^{\varepsilon}(\mathrm{d}x) &= \int_{\mathbb{R}^d} \frac{1}{2\varepsilon^2}|Qx|^2 \bar{\mu}^{\varepsilon}(\mathrm{d}x) + \mathcal{O}\bigl(\varepsilon^{4\nu-2}\bigr) \\ &= \varepsilon^{2\nu-2}\Lambda^+(M) + \mathcal{O}\bigl(\varepsilon^{4\nu-2}\bigr). \end{aligned} \tag{3.14}$$

To prove the second equation in (3.11), we use the bound

$$\bigl|\ell(x) - \ell(z) - D\ell(z)(x-z)\bigr| \le \tilde{C}_{\ell}|x-z|^2 \qquad \forall x \in \mathbb{R}^d, \forall z \in \mathcal{S}, \tag{3.15}$$

for some constant $\tilde{C}_\ell$, and since $\bar{\mu}^\varepsilon$ has zero mean, we obtain

$$\left|\int_{\mathbb{R}^d}\bigl(\ell(x)-\ell(0)\bigr)\bar{\mu}^\varepsilon(\mathrm{d}x)\right| \le \varepsilon^{2\nu}\tilde{C}_\ell \operatorname{trace}(\Sigma). \tag{3.16}$$

By combining (3.14) and (3.16), we obtain (3.11). The proof is complete. □

Recall the notation in Definition 1.10. Lemma 3.2 in conjunction with Lemma 3.1 leads to the following.

LEMMA 3.3. *It holds that*

$$\begin{aligned}\beta^\varepsilon_* &\le \mathfrak{J} + \varepsilon^{2\nu-2}\min_{z\in\mathcal{Z}} \Lambda^+\bigl(Dm(z)\bigr) + \mathcal{O}\bigl(\varepsilon^{2\nu}\bigr) \qquad \text{if } \nu>1,\\ \beta^\varepsilon_* &\le \mathfrak{J}_{\mathrm{c}} + \mathcal{O}\bigl(\varepsilon^2\bigr) \qquad \text{if } \nu=1.\end{aligned} \tag{3.17}$$

Moreover, if $\nu>1$, *then*

$$\lim_{\varepsilon\searrow 0}\beta^\varepsilon_* = \mathfrak{J}, \tag{3.18}$$

and $\mathfrak{S}\subset\mathcal{Z}$.

PROOF. Recall the function $\mathcal{R}[v]$ defined in (3.1). By Lemma 3.2, we have

$$\begin{aligned}\beta^\varepsilon_* &\le \int_{\mathbb{R}^d}\mathcal{R}\bigl[\bar{v}^\varepsilon_z\bigr](x)\bar{\mu}^\varepsilon_z(\mathrm{d}x)\\ &\le \ell(z)+\varepsilon^{2\nu-2}\Lambda^+\bigl(Dm(z)\bigr)+\mathcal{O}\bigl(\varepsilon^{2\nu}\bigr) \qquad \forall z\in\mathcal{S},\ \nu\ge 1.\end{aligned} \tag{3.19}$$

Since $\ell(z)=\mathfrak{J}$ for all $z\in\widetilde{\mathcal{Z}}\subset\mathcal{Z}$, the first and the second inequalities in (3.17) follow by evaluating (3.19) at a point $z\in\widetilde{\mathcal{Z}}$, and $z\in\mathcal{Z}_{\mathrm{c}}$, respectively.

Since

$$\lim_{\varepsilon\searrow 0}\beta^\varepsilon_* \ge \mathfrak{J} \tag{3.20}$$

for all $\nu>0$ by Lemma 3.1, equation (3.18) follows by (3.17) and (3.20) when $\nu>1$, and clearly then, in this case we have $\mathfrak{S}\subset\mathcal{Z}$. □

REMARK 3.4. It is worth mentioning here that if $z\in\mathcal{S}_{\mathrm{s}}$, then a control that renders $\{z\}$ stochastically stable can be synthesized from the energy function $\mathcal{V}$. Note that by Theorem 2.2(ii), $\mathcal{V}$ can be selected so that $\mathcal{V}(z)=0$ and $\mathcal{V}(z')>0$ for all $z'\in\mathcal{S}\setminus\{z\}$. Consider the control

$$\check{v}^\varepsilon(x) := -\frac{1}{\varepsilon}\bigl(m(x)+\nabla\mathcal{V}(x)\bigr), \qquad t\ge 0.$$

Then X is given by

$$\mathrm{d}X_t = -\nabla\mathcal{V}(X_t)\,\mathrm{d}t + \varepsilon^\nu\,\mathrm{d}W_t, \qquad t\ge 0.$$

Let $\check{\mu}^\varepsilon$ denote its unique invariant probability measure. Recall the definition in (1.15). Since

$$\mathcal{L}^\varepsilon_{\check{v}^\varepsilon}\mathcal{V} \le \frac{\varepsilon^{2\nu}}{2}\|\Delta\mathcal{V}\|_\infty - |\nabla\mathcal{V}|^2,$$

it follows that

$$2\int_{\mathbb{R}^d} |\nabla\mathcal{V}|^2 \,\mathrm{d}\check{\mu}^\varepsilon \le \varepsilon^{2\nu}\|\Delta\mathcal{V}\|_\infty.$$

Note that $\check{\mu}^\varepsilon$ has density $\varrho^\varepsilon(x) = C(\varepsilon)\mathrm{e}^{-\frac{2\mathcal{V}(x)}{\varepsilon^{2\nu}}}$, where $C(\varepsilon)$ is a normalizing constant. Thus, we obtain

$$\begin{aligned}\int_{\mathbb{R}^d} \bigl|\check{v}^\varepsilon(x)\bigr|^2\check{\mu}^\varepsilon(\mathrm{d}x) &\le 2\int_{\mathbb{R}^d} \bigl(\bigl|m(x)\bigr|^2 + \bigl|\nabla\mathcal{V}(x)\bigr|^2\bigr)\varepsilon^{-2}\check{\mu}^\varepsilon(\mathrm{d}x)\\ &\le 2\int_{\mathbb{R}^d} \varepsilon^{-2}\bigl|m(x)\bigr|^2\check{\mu}^\varepsilon(\mathrm{d}x) + \varepsilon^{2\nu-2}\|\Delta\mathcal{V}\|_\infty\\ &\le \mathcal{O}\bigl(\varepsilon^{2\nu-2}\bigr) + \varepsilon^{2\nu-2}\|\Delta\mathcal{V}\|_\infty.\end{aligned}$$

For the last inequality, we use the fact that m is bounded, $m(z) = 0$, and that $\mathcal{V}$ is locally quadratic around z.

3.2. *Results concerning stable equilibria.* Recall that $\mathcal{S}_{\mathrm{s}}$ is the collection of stable equilibrium points, and $\mathfrak{J}_{\mathrm{s}} = \min_{z\in\mathcal{S}_{\mathrm{s}}}\{\ell(z)\}$. The following lemma holds for any $\nu > 0$. It shows that if $z \in \mathcal{S}_{\mathrm{s}}$, then there exists a Markov stationary control v^ε with invariant measure μ^ε satisfying $\int_{\mathbb{R}^d} |v^\varepsilon(x)|^2\mu^\varepsilon(\mathrm{d}x) \in \mathcal{O}(\varepsilon^n)$ for any $n \in \mathbb{N}$, under which $\{z\}$ is stochastically stable.

LEMMA 3.5. *The following hold.*

(i) *For any $\nu > 0$ and $z \in \mathcal{S}_{\mathrm{s}}$ there exists a Markov control $\check{v}^\varepsilon$, and constants $\varepsilon_0 = \varepsilon_0(\nu) > 0$, and $c_0 > 0$ independent of ν, with the following properties. With $\check{\mu}^\varepsilon$ denoting the invariant probability measure of (1.1) under the control $\check{v}^\varepsilon$, it holds that*

$$\begin{aligned}\int_{|x-z|\ge\varepsilon^{\frac{\nu}{2}}} |x-z|^2\check{\mu}^\varepsilon(\mathrm{d}x) &\le \frac{\varepsilon^{2\nu}}{c_0(1-\varepsilon^\nu)}\mathrm{e}^{-c_0\varepsilon^{-\nu}},\\ \int_{\mathbb{R}^d} \bigl|\check{v}^\varepsilon(x)\bigr|^2\check{\mu}^\varepsilon(\mathrm{d}x) &\le \frac{\varepsilon^{2(\nu-1)}}{c_0(1-\varepsilon^\nu)}\mathrm{e}^{-c_0\varepsilon^{-\nu}},\end{aligned} \tag{3.21}$$

for all $\varepsilon < \varepsilon_0$, and

$$\varepsilon^{-\nu}\left|\int_{\mathbb{R}^d} \ell(x)\check{\mu}^\varepsilon(\mathrm{d}x) - \ell(z)\right| \xrightarrow[\varepsilon\searrow 0]{} 0. \tag{3.22}$$

In particular, we have

$$\limsup_{\varepsilon\searrow 0} \frac{1}{\varepsilon^n}\int_{\mathbb{R}^d} \bigl|\check{v}^\varepsilon(x)\bigr|^2\check{\mu}^\varepsilon(\mathrm{d}x) = 0 \qquad \forall n \in \mathbb{N}.$$

(ii) *It holds that* $\beta_*^\varepsilon \le \mathfrak{J}_{\mathrm{s}} + \mathfrak{o}(\varepsilon^\nu)$ *for* $\nu \in (0, \frac{2}{3})$, *and* $\beta_*^\varepsilon \le \mathfrak{J}_{\mathrm{s}} + \mathcal{O}(\varepsilon^{4\nu-2})$ *for* $\nu \in [\frac{2}{3}, 1)$.

PROOF. In order to simplify the notation, we translate the origin so that $z = 0$, and we let $M := Dm(0)$. Let R^{-1} be the symmetric positive definite solution to the Lyapunov equation $MR^{-1} + R^{-1}M^{\mathsf{T}} = -4I$. Thus, $M^{\mathsf{T}}R + RM = -4R^2$. Since scaling R by multiplying it with a positive constant smaller than 1 preserves the inequality

$$M^{\mathsf{T}}R + RM \le -4R^2, \tag{3.23}$$

we may assume that $\operatorname{trace}(R) \le 1$ and (3.23) holds. The sole purpose of this scaling is to simplify the calculations in the proof. We define the control $\check{v}^\varepsilon$ by

$$\check{v}^\varepsilon(x) := \begin{cases} \varepsilon^{-1}\big(Mx - m(x)\big) & \text{if } |Rx| \ge \varepsilon^{\frac{\nu}{2}}, \\ 0 & \text{otherwise.} \end{cases}$$

We apply the function $F(x) := \varepsilon^{2\nu}\exp(\varepsilon^{-2\nu}\langle x, Rx\rangle)$ to $\mathcal{L}^\varepsilon_{\check{v}^\varepsilon}$, which is defined in (1.15). By (3.23), and since $\operatorname{trace}(R) \le 1$, we obtain

$$\begin{aligned} \mathcal{L}^\varepsilon_{\check{v}^\varepsilon} F(x) &= \big(\varepsilon^{2\nu}\operatorname{trace}(R) + 2|Rx|^2 + \langle x, (M^{\mathsf{T}}R + RM)x\rangle\big)\mathrm{e}^{\frac{\langle x, Rx\rangle}{\varepsilon^{2\nu}}} \\ &\le \big(\varepsilon^{2\nu} - 2|Rx|^2\big)\mathrm{e}^{\frac{\langle x, Rx\rangle}{\varepsilon^{2\nu}}} \qquad \text{if } |Rx| \ge \varepsilon^{\frac{\nu}{2}}. \end{aligned} \tag{3.24}$$

If $|Rx| < \varepsilon^{\frac{\nu}{2}}$, then $\check{v}^\varepsilon = 0$, and we obtain

$$\begin{aligned} \mathcal{L}^\varepsilon_{\check{v}^\varepsilon} F(x) &= \big(\varepsilon^{2\nu}\operatorname{trace}(R) + 2|Rx|^2 + 2\langle m(x), Rx\rangle\big)\mathrm{e}^{\frac{\langle x, Rx\rangle}{\varepsilon^{2\nu}}} \\ &\le \big(\varepsilon^{2\nu} - 2|Rx|^2 + 2|Mx - m(x)||Rx|\big)\mathrm{e}^{\frac{\langle x, Rx\rangle}{\varepsilon^{2\nu}}} \\ &\le \big(\varepsilon^{2\nu} - |Rx|^2\big)\mathrm{e}^{\frac{\langle x, Rx\rangle}{\varepsilon^{2\nu}}} \end{aligned} \tag{3.25}$$

provided that $|Rx| < \varepsilon^{\frac{\nu}{2}} \wedge \frac{1}{2}\|R^{-1}\|^{-2}\tilde{C}_m^{-1}$, where in the first inequality we use (3.23), and in the second we use (2.2). Thus, selecting ε_0 as

$$\varepsilon_0 := 1 \wedge \left(\frac{1}{2}\|R^{-1}\|^{-2}\tilde{C}_m^{-1}\right)^{\frac{2}{\nu}},$$

then, provided that $\varepsilon < \varepsilon_0$, (3.25) holds for all x such that $|Rx| < \varepsilon^{\frac{\nu}{2}}$. It follows by (3.24) and (3.25) that $\mathcal{L}^\varepsilon_{\check{v}^\varepsilon} F(x) \le 0$ if $|Rx| \ge \varepsilon^\nu$, and

$$\sup\big\{\mathcal{L}^\varepsilon_{\check{v}^\varepsilon} F(x) \colon |Rx| \le \varepsilon^\nu,\ \varepsilon < \varepsilon_0\big\} \le \mathrm{e}^{\|R^{-1}\|}\varepsilon^{2\nu} \qquad \forall \varepsilon < \varepsilon_0. \tag{3.26}$$

Thus, by (3.24), (3.25), and (3.26), we obtain

$$\mathcal{L}^\varepsilon_{\check{v}^\varepsilon} F(x) \le \mathrm{e}^{\|R^{-1}\|}\varepsilon^{2\nu}\mathbb{1}_{\{|Rx| \le \varepsilon^\nu\}} - \big(|Rx|^2 - \varepsilon^{2\nu}\big)\mathrm{e}^{\frac{\langle x, Rx\rangle}{\varepsilon^{2\nu}}}\mathbb{1}_{\{|Rx| \ge \varepsilon^\nu\}} \tag{3.27}$$

for all $x \in \mathbb{R}^d$ and $\varepsilon < \varepsilon_0$. Note that (3.27) is a Foster–Lyapunov equation and F is inf-compact. It follows that $\check{v}^\varepsilon$ is a stable Markov control with invariant measure $\check{\mu}^\varepsilon$. Thus, integrating (3.27) with respect to the invariant probability measure $\check{\mu}^\varepsilon$, we obtain

$$\int_{\{|Rx| \ge \varepsilon^\nu\}} \bigl(|Rx|^2 - \varepsilon^{2\nu}\bigr) \mathrm{e}^{\frac{\langle x, Rx\rangle}{\varepsilon^{2\nu}}} \check{\mu}^\varepsilon(\mathrm{d}x) \le \mathrm{e}^{\|R^{-1}\|} \varepsilon^{2\nu} \qquad \forall \varepsilon < \varepsilon_0. \tag{3.28}$$

For any $a \in (0, 1)$, we have

$$|y|^2 \le \frac{|y|^2 - a^4}{1 - a^2} \qquad \text{if } |y| \ge a. \tag{3.29}$$

Thus, using (3.28), and applying (3.29) with $a = \varepsilon^{\frac{\nu}{2}}$, and the inequality $\langle x, Rx\rangle \ge \|R\|^{-1} |Rx|^2$, we obtain

$$\begin{aligned} &\int_{|Rx| \ge \varepsilon^{\frac{\nu}{2}}} |Rx|^2 \check{\mu}^\varepsilon(\mathrm{d}x) \\ &\quad \le \int_{|Rx| \ge \varepsilon^{\frac{\nu}{2}}} \frac{|Rx|^2 - \varepsilon^{2\nu}}{1 - \varepsilon^\nu} \mathrm{e}^{-\|R\|^{-1}\varepsilon^{-\nu}} \mathrm{e}^{\frac{\langle x, Rx\rangle}{\varepsilon^{2\nu}}} \check{\mu}^\varepsilon(\mathrm{d}x) \\ &\quad \le \frac{1}{1 - \varepsilon^\nu} \mathrm{e}^{-\|R\|^{-1}\varepsilon^{-\nu}} \int_{|Rx| \ge \varepsilon^\nu} \bigl(|Rx|^2 - \varepsilon^{2\nu}\bigr) \mathrm{e}^{\frac{\langle x, Rx\rangle}{\varepsilon^{2\nu}}} \check{\mu}^\varepsilon(\mathrm{d}x) \\ &\quad \le \mathrm{e}^{\|R^{-1}\|} \frac{\varepsilon^{2\nu}}{1 - \varepsilon^\nu} \mathrm{e}^{-\|R\|^{-1}\varepsilon^{-\nu}} \qquad \forall \varepsilon < \varepsilon_0. \end{aligned} \tag{3.30}$$

Similarly, by (3.28), and using the inequality

$$\bigl(N^2 - 1\bigr)|y|^2 \le N^2\bigl(|y|^2 - \varepsilon^{2\nu}\bigr) \qquad \text{for } |y| \ge N\varepsilon^\nu,$$

and for any $N \ge 2$, we obtain

$$\int_{|Rx| \ge N\varepsilon^\nu} |Rx|^2 \check{\mu}^\varepsilon(\mathrm{d}x) \le \mathrm{e}^{\|R^{-1}\|} \frac{N^2 \varepsilon^{2\nu}}{N^2 - 1} \mathrm{e}^{-N^{-2}\varepsilon^{-2\nu}\|R\|^{-1}} \tag{3.31}$$

for all $\varepsilon < \varepsilon_0$.

In addition, since $\check{v}^\varepsilon = 0$ for $|Rx| \le \varepsilon^{\frac{\nu}{2}}$ by definition, and $|\check{v}^\varepsilon(x)| \le \tilde{C}_m \frac{|x|}{\varepsilon}$ by (2.2), it follows by (3.30) that

$$\int_{\mathbb{R}^d} \bigl|\check{v}^\varepsilon(x)\bigr|^2 \check{\mu}^\varepsilon(\mathrm{d}x) \le \bigl\|R^{-1}\bigr\|^2 \frac{\tilde{C}_m^2}{1 - \varepsilon^\nu} \mathrm{e}^{\|R^{-1}\|} \varepsilon^{2\nu - 2} \mathrm{e}^{-\|R\|^{-1}\varepsilon^{-\nu}} \tag{3.32}$$

for all $\varepsilon < \varepsilon_0$. Then (3.21) follows from (3.30) and (3.32), by choosing a common constant c_0.

Consider the "scaled" diffusion

$$\mathrm{d}\hat{X}_t = \hat{b}^\varepsilon(\hat{X}_t)\,\mathrm{d}t + \mathrm{d}W_t, \qquad t \ge 0,$$

where

$$\hat{b}^\varepsilon := \frac{m(\varepsilon^\nu x) + \varepsilon \check{v}^\varepsilon(\varepsilon^\nu x)}{\varepsilon^\nu},$$

and let $\hat{\mu}^\varepsilon$ denote its invariant probability measure. With $\check{\varrho}^\varepsilon$ and $\hat{\varrho}^\varepsilon$ denoting the densities of $\check{\mu}^\varepsilon$ and $\hat{\mu}^\varepsilon$, respectively, we have $\varepsilon^{\nu d}\check{\varrho}^\varepsilon(\varepsilon^\nu x) = \hat{\varrho}^\varepsilon(x)$ for all $x \in \mathbb{R}^d$. Substituting $x = \varepsilon^\nu y$ in (3.28), we deduce that the family of probability measures $\{\hat{\mu}^\varepsilon : \varepsilon \in (0,1)\}$ is tight. At the same time, the (discontinuous) drift $\hat{b}^\varepsilon$ converges to Mx as $\varepsilon \searrow 0$, uniformly on compact sets. We claim that $\hat{\varrho}^\varepsilon$ converges, as $\varepsilon \searrow 0$, to the Gaussian density ρ_Σ with mean 0 and covariance matrix Σ, given by $M\Sigma + \Sigma M^\mathsf{T} = -I$, that is, $\Sigma = \frac{1}{4}R^{-1}$, uniformly on compact sets. Indeed, since $\hat{b}^\varepsilon$ is locally bounded uniformly in $\varepsilon \in (0,1)$, and the family $\{\hat{\mu}^\varepsilon, \varepsilon \in (0,1)\}$ is tight, the densities $\hat{\varrho}^\varepsilon$ of $\hat{\mu}^\varepsilon$ are locally Hölder equicontinuous [see Lemma 3.2.4 in Arapostathis, Borkar and Ghosh (2012)]. Let $\hat{\varrho}$ be any limit point of $\hat{\varrho}^{\varepsilon_n}$ in $\mathcal{C}(\mathbb{R}^d)$ along some sequence $\varepsilon_n \searrow 0$. Since $\{\hat{\mu}^\varepsilon : \varepsilon \in (0,1)\}$ is tight, it follows that $\hat{\varrho}^{\varepsilon_n}$ also converges in $L^1(\mathbb{R}^d)$, as $n \to \infty$, and hence $\int_{\mathbb{R}^d} \hat{\varrho}(x)\,\mathrm{d}x = 1$. With

$$\hat{\mathcal{L}}^\varepsilon := \frac{1}{2}\Delta + \langle \hat{b}^\varepsilon, \nabla \rangle \quad \text{and} \quad \hat{\mathcal{L}}^0 := \frac{1}{2}\Delta + \langle Mx, \nabla \rangle,$$

and since $\int_{\mathbb{R}^d} \hat{\mathcal{L}}^\varepsilon f(x)\hat{\varrho}^\varepsilon(x)\,\mathrm{d}x = 0$ for all $f \in \mathcal{C}_\mathrm{c}^\infty(\mathbb{R}^d)$, we have

$$\begin{aligned}\int_{\mathbb{R}^d} \hat{\mathcal{L}}^0 f(x)\hat{\varrho}(x)\,\mathrm{d}x &= \int_{\mathbb{R}^d} \big(\hat{\mathcal{L}}^0 f(x) - \hat{\mathcal{L}}^\varepsilon f(x)\big)\hat{\varrho}(x)\,\mathrm{d}x \\ &\quad + \int_{\mathbb{R}^d} \hat{\mathcal{L}}^\varepsilon f(x)\big(\hat{\varrho}(x) - \hat{\varrho}^\varepsilon(x)\big)\,\mathrm{d}x\end{aligned} \tag{3.33}$$

for all $f \in \mathcal{C}_\mathrm{c}^\infty(\mathbb{R}^d)$. It is clear that both terms on the right-hand side of (3.33) converge to 0 as $\varepsilon = \varepsilon_n \searrow 0$. This implies that $\hat{\varrho}$ is the density of the invariant probability measure of the diffusion $\mathrm{d}X_t = MX_t\,\mathrm{d}t + \mathrm{d}W_t$, which is Gaussian as claimed.

Since the Gaussian density ρ_Σ has zero mean, then by the uniform integrability property, implied by (3.31), we have

$$\varepsilon^{-\nu} \int_{\mathbb{R}^d} \big(D\ell(0)x\big)\check{\mu}^\varepsilon(\mathrm{d}x) \xrightarrow[\varepsilon \searrow 0]{} 0. \tag{3.34}$$

On the other hand, it follows by (3.31) that for some constant $\kappa_1 > 0$ we have $\int_{\mathbb{R}^d} |x|^2 \check{\mu}^\varepsilon(\mathrm{d}x) < \kappa_1 \varepsilon^{2\nu}$ for all $\varepsilon < \varepsilon_0$. Thus, using (3.15), we obtain

$$\varepsilon^{-\nu} \int_{\mathbb{R}^d} \big|\ell(x) - \ell(0) - D\ell(0)x\big|\check{\mu}^\varepsilon(\mathrm{d}x) \le \kappa_1 \tilde{C}_\ell \varepsilon^\nu. \tag{3.35}$$

Combining (3.34)–(3.35), we obtain (3.22).

Next, we turn to part (ii). Consider the control $v^\varepsilon(x) = \varepsilon^{-1}(Mx - m(x))$ for $x \in \mathbb{R}^d$. Then $m(x) + \varepsilon v^\varepsilon(x) = Mx$, and the associated invariant measure μ^ε is

Gaussian with mean 0 and covariance matrix $\varepsilon^{2\nu}\Sigma$. Using the bound in (2.2), we obtain

$$\int_{\mathbb{R}^d} |v^\varepsilon|^2 \, \mathrm{d}\mu^\varepsilon \le \int_{\mathbb{R}^d} \tilde{C}_m^2 \varepsilon^{-2} |x|^4 \mu^\varepsilon(\mathrm{d}x) \in \mathcal{O}\big(\varepsilon^{4\nu-2}\big).$$

Since μ^ε has zero mean, using a triangle inequality and (3.15), as in the proof of Lemma 3.2, we obtain

$$\left| \int_{\mathbb{R}^d} \big(\ell(x) - \ell(0)\big) \mu^\varepsilon(\mathrm{d}x) \right| \le \varepsilon^{2\nu} \tilde{C}_\ell \, \mathrm{trace}(\Sigma).$$

Since $4\nu - 2 < 2\nu$ for $\nu < 1$, we obtain $\beta_*^\varepsilon \le \mathfrak{J}_{\mathrm{s}} + \mathcal{O}(\varepsilon^{4\nu-2})$. On the other hand, by part (1) we already know that $\beta_*^\varepsilon \le \mathfrak{J}_{\mathrm{s}} + \mathfrak{o}(\varepsilon^\nu)$. To complete the proof, we observe that $\nu \le 4\nu - 2$ for $\nu \ge \frac{2}{3}$, and $\nu > 4\nu - 2$ for $\nu < \frac{2}{3}$. □

3.3. *Results concerning the subcritical regime.* By Lemma 3.5, we can always find a stable admissible control such that the corresponding invariant probability measure concentrates on a stable equilibrium point as $\varepsilon \searrow 0$, while keeping the ergodic cost in (1.3) bounded, uniformly in $\varepsilon \in (0, 1)$. Now we proceed to show that for $\nu < 1$, η_*^ε concentrates on $\mathcal{S}_{\mathrm{s}}$.

LEMMA 3.6. *Suppose* $\nu < 1$. *Then*

$$\eta_*^\varepsilon(\mathcal{S} \setminus \mathcal{S}_{\mathrm{s}}) \xrightarrow[\varepsilon \searrow 0]{} 0 \quad \textit{and} \quad \lim_{\varepsilon \searrow 0} \beta_*^\varepsilon = \mathfrak{J}_{\mathrm{s}}.$$

PROOF. We argue by contradiction. Suppose that

$$\limsup_{\varepsilon \searrow 0} \eta_*^\varepsilon\big(B_r(z)\big) > 0$$

for some $r > 0$ and $z \notin \mathcal{S}_{\mathrm{s}}$. In Theorem 2.2, we may select a_z such that $a_z \ne a_{z'}$ for $z \ne z'$. Thus, by Theorem 2.2(ii), there exists $\delta > 0$ such that the interval $(\mathcal{V}(z) - 3\delta, \mathcal{V}(z) + 3\delta)$ contains no other critical values of $\mathcal{V}$ other than $\mathcal{V}(z)$. Let $\varphi \in \mathcal{C}^2(\mathbb{R})$ be such that:

(a) $\varphi(\mathcal{V}(z) + y) = y$ for $y \in (\mathcal{V}(z) - \delta, \mathcal{V}(z) + \delta)$;
(b) $\varphi' \in [0, 1]$ on $(\mathcal{V}(z) - 2\delta, \mathcal{V}(z) + 2\delta)$;
(c) $\varphi' = 0$ on $(\mathcal{V}(z) - 2\delta, \mathcal{V}(z) + 2\delta)^c$.

Select $r > 0$ such that

$$\sup_{x \in B_r(z)} \big|\Delta\mathcal{V}(x) - \Delta\mathcal{V}(z)\big| < \frac{1}{2} \big|\Delta\mathcal{V}(z)\big|. \tag{3.36}$$

Note that by Theorem 2.2 and Lemma 2.3, the function $\mathcal{V}$ takes distinct values on $\mathcal{S}$. Therefore, we may also choose the radius r small enough so that

$$B_r(z) \subset \big\{x : \big|\mathcal{V}(x) - \mathcal{V}(z)\big| \le \delta\big\} \subset B_r^c\big(S \setminus \{z\}\big).$$

By the infinitesimal characterization of an invariant probability measure we have

$$\int_{\mathbb{R}^d} \mathcal{L}^{\varepsilon}_{v^{\varepsilon}_*}(\varphi\circ\mathcal{V})(x)\eta^{\varepsilon}_*(\mathrm{d}x)=0,$$

which we write as

$$\begin{aligned}&\frac{\varepsilon^{2\nu}}{2}\biggl(\int_{\mathbb{R}^d}\varphi'(\mathcal{V})\Delta\mathcal{V}\,\mathrm{d}\eta^{\varepsilon}_*+\int_{\mathbb{R}^d}\varphi''(\mathcal{V})|\nabla\mathcal{V}|^2\,\mathrm{d}\eta^{\varepsilon}_*\biggr)\\&\qquad+\varepsilon\int_{\mathbb{R}^d}\varphi'(\mathcal{V})\langle v^{\varepsilon}_*,\nabla\mathcal{V}\rangle\,\mathrm{d}\eta^{\varepsilon}_*+\int_{\mathbb{R}^d}\varphi'(\mathcal{V})\langle m,\nabla\mathcal{V}\rangle\,\mathrm{d}\eta^{\varepsilon}_*=0.\end{aligned}\tag{3.37}$$

Recall the definition of the optimal control effort $\mathcal{G}^{\varepsilon}_*$ in (1.18), and also define

$$\begin{aligned}&\zeta^{\varepsilon}:=\biggl(\int_{\mathbb{R}^d}\varphi'(\mathcal{V})|\nabla\mathcal{V}|^2\,\mathrm{d}\eta^{\varepsilon}_*\biggr)^{\frac{1}{2}},\\&\xi^{\varepsilon}_1:=\frac{1}{2}\int_{\mathbb{R}^d}\varphi'(\mathcal{V})\Delta\mathcal{V}\,\mathrm{d}\eta^{\varepsilon}_*,\qquad\xi^{\varepsilon}_2:=\frac{1}{2}\int_{\mathbb{R}^d}\varphi''(\mathcal{V})|\nabla\mathcal{V}|^2\,\mathrm{d}\eta^{\varepsilon}_*,\end{aligned}\tag{3.38}$$

and $\xi^{\varepsilon}:=\xi^{\varepsilon}_1+\xi^{\varepsilon}_2$. By the Cauchy–Schwarz inequality we have

$$\biggl|\int_{\mathbb{R}^d}\varphi'(\mathcal{V})\langle v^{\varepsilon}_*,\nabla\mathcal{V}\rangle\,\mathrm{d}\eta^{\varepsilon}_*\biggr|\le\|\sqrt{\varphi'}\|_{\infty}\sqrt{2\mathcal{G}^{\varepsilon}_*}\zeta^{\varepsilon}\le\sqrt{2\mathcal{G}^{\varepsilon}_*}\zeta^{\varepsilon}.\tag{3.39}$$

By Theorem 2.2(iv), we have $C_0(\zeta^{\varepsilon})^2\le-\int_{\mathbb{R}^d}\varphi'(\mathcal{V})\langle m,\nabla\mathcal{V}\rangle\,\mathrm{d}\eta^{\varepsilon}_*$. Therefore, by (3.37) and (3.39) we obtain

$$C_0\bigl(\zeta^{\varepsilon}\bigr)^2-\varepsilon\sqrt{2\mathcal{G}^{\varepsilon}_*}\zeta^{\varepsilon}-\varepsilon^{2\nu}\xi^{\varepsilon}\le0.\tag{3.40}$$

We write

$$\xi^{\varepsilon}_1=\int_{B_r(z)}\varphi'(\mathcal{V})\Delta\mathcal{V}\,\mathrm{d}\eta^{\varepsilon}_*+\int_{B^c_r(z)}\varphi'(\mathcal{V})\Delta\mathcal{V}\,\mathrm{d}\eta^{\varepsilon}_*.\tag{3.41}$$

Since $\mathcal{V}$ is inf-compact, it follows that $\varphi\circ\mathcal{V}$ is constant outside a compact set. Therefore, the support of $\varphi'(\mathcal{V}(\cdot))$ is compact, and as a result $\Delta\mathcal{V}$ is bounded on this set. By (3.36), (3.41), Theorem 2.2(iii), and since $\eta^{\varepsilon}_*(B^c_r(\mathcal{S}))\searrow0$ as $\varepsilon\searrow0$ (by Lemma 3.1), we obtain

$$\limsup_{\varepsilon\searrow0}\bigl(-\xi^{\varepsilon}_1\bigr)\ge-\frac{1}{2}\Delta\mathcal{V}(z)\limsup_{\varepsilon\searrow0}\eta^{\varepsilon}_*\bigl(B_r(z)\bigr)>0.\tag{3.42}$$

On the other hand, since $\varphi''(\mathcal{V})=0$ on some open neighborhood of $\mathcal{S}$, it follows that $\xi^{\varepsilon}_2\to0$ as $\varepsilon\searrow0$. Therefore, we have $\limsup_{\varepsilon\searrow0}(-\xi^{\varepsilon})>0$. However, since the discriminant of (3.40) must be nonnegative, we obtain

$$\varepsilon^2\mathcal{G}^{\varepsilon}_*\ge-2C_0\varepsilon^{2\nu}\xi^{\varepsilon},\tag{3.43}$$

which leads to a contradiction. Hence, $\eta^\varepsilon_*(\mathcal{S} \setminus \mathcal{S}_{\rm s}) \xrightarrow[\varepsilon \searrow 0]{} 0$. This implies that $\liminf_{\varepsilon \searrow 0} \beta^\varepsilon_* \ge \mathfrak{J}_{\rm s}$, which combined with Lemma 3.5(ii), establishes the second claim of the lemma. □

We revisit the subcritical regime in Corollary 4.2 to obtain a lower bound for β^ε_*.

It is worthwhile to present at this point the following one-dimensional example, which shows how the value of β^ε_* for small ε bifurcates as we cross the critical regime.

EXAMPLE 3.7. Let $d = 1$, $m(x) = Mx$, and $\ell(x) = \frac{1}{2}Lx^2$, with $M > 0$ and $L > 0$. Then the solution to (1.16) takes the form

$$V^\varepsilon = \frac{M + \sqrt{M^2 + L\varepsilon^2}}{2\varepsilon^2} x^2,$$

$$\beta^\varepsilon_* = \frac{\varepsilon^{2\nu-2}}{2}\Big(M + \sqrt{M^2 + L\varepsilon^2}\Big).$$

Note that $\beta^\varepsilon_* \to \ell(0) = 0$, $\beta^\varepsilon_* \to M$, and $\beta^\varepsilon_* \to \infty$, as $\varepsilon \searrow 0$, when $\nu > 1$, $\nu = 1$, and $\nu < 1$, respectively.

4. Concentration bounds for the optimal stationary distribution. We start with the following lemma, which is valid for all ν.

LEMMA 4.1. *For any bounded domain G there exists a constant $\hat\kappa_0 = \hat\kappa_0(G, \nu)$ such that*

$$\int_G \big(\operatorname{dist}(x, \mathcal{S})\big)^2 \eta^\varepsilon_*(\mathrm{d}x) \le \hat\kappa_0 \varepsilon^{2(\nu \wedge 2)} \qquad \forall \nu > 0, \forall \varepsilon \in (0, 1). \tag{4.1}$$

PROOF. We fix some bounded domain G which without loss of generality contains $\mathcal{S}$, and choose some number δ such that $\delta \ge \sup_{x \in G} \mathcal{V}(x)$. Without loss of generality, we assume that $\ell(x) > \mathfrak{J}$ for all $x \in G^c$; otherwise we enlarge G. Let $\tilde\varphi : \mathbb{R} \to \mathbb{R}$ be a smooth function such that:

(a) $\tilde\varphi(y) = y$ for $y \in (-\infty, \delta)$;
(b) $\tilde\varphi' \in (0, 1)$ on $(\delta, 2\delta)$;
(c) $\tilde\varphi' = 0$ on $[2\delta, \infty)$;
(d) $\tilde\varphi'' \le 0$.

Define $\tilde\zeta^\varepsilon$, $\tilde\xi^\varepsilon_1$, and $\tilde\xi^\varepsilon_2$, as in (3.38) by replacing φ with $\tilde\varphi$, and let $\tilde\xi^\varepsilon := \tilde\xi^\varepsilon_1 + \tilde\xi^\varepsilon_2$. As in (3.40) we obtain

$$C_0\big(\tilde\zeta^\varepsilon\big)^2 - \varepsilon\sqrt{2\mathcal{G}^\varepsilon_*}\tilde\zeta^\varepsilon - \varepsilon^{2\nu}\tilde\xi^\varepsilon \le 0. \tag{4.2}$$

By Theorem 2.2(iv), we have

$$\int_{\{x\colon \mathcal{V}(x)\le\delta\}} \big(\operatorname{dist}(x,\mathcal{S})\big)^2 \eta_*^\varepsilon(\mathrm{d}x) \le C_0^{-1}(\tilde\zeta^\varepsilon)^2. \tag{4.3}$$

By an application of Young's inequality to (4.2), we obtain

$$\frac{C_0}{2}(\tilde\zeta^\varepsilon)^2 - \frac{1}{C_0}\varepsilon^2 \mathcal{G}_*^\varepsilon - \varepsilon^{2\nu}\tilde\xi^\varepsilon \le 0,$$

and hence we have $\tilde\zeta^\varepsilon \in \mathcal{O}(\varepsilon^{\nu\wedge 2})$. Thus (4.1) follows by (4.3). □

COROLLARY 4.2. *Suppose $\nu \ge 1$. Then following hold.*

(a) *The optimal control effort $\mathcal{G}_*^\varepsilon$ satisfies*

$$\begin{aligned} &\mathcal{G}_*^\varepsilon \in \mathcal{O}\big(\varepsilon^{\nu\wedge 2}\big) \qquad \text{if } \mathfrak{J} = \mathfrak{J}_{\mathrm{s}},\ \nu > 1,\ \text{or if } \mathfrak{J}_{\mathrm{c}} = \mathfrak{J}_{\mathrm{s}},\ \nu = 1,\\ &\mathcal{G}_*^\varepsilon \in \mathcal{O}\big(\varepsilon^{(2\nu-2)\wedge 2}\big) \qquad \text{if } \mathfrak{J} < \mathfrak{J}_{\mathrm{s}} \text{ and } \nu > 1 \end{aligned} \tag{4.4}$$

and

$$\liminf_{\varepsilon\searrow 0} \frac{1}{\varepsilon^{2\nu-2}} \mathcal{G}_*^\varepsilon > 0 \qquad \text{if } \mathfrak{J} < \mathfrak{J}_{\mathrm{s}} \text{ and } \nu > 1. \tag{4.5}$$

(b) $\beta_*^\varepsilon - \mathfrak{J} \ge \mathcal{O}(\varepsilon^{\nu\wedge 2})$ *for* $\nu > 1$.

PROOF. Select a domain G as in the proof of Lemma 4.1. Define $\tilde\zeta^\varepsilon$, $\tilde\xi_1^\varepsilon$, and $\tilde\xi_2^\varepsilon$ as in (3.38) by replacing φ with $\tilde\varphi$, and let $\tilde\xi^\varepsilon := \tilde\xi_1^\varepsilon + \tilde\xi_2^\varepsilon$. Then (4.2) holds, and thus $\tilde\zeta^\varepsilon \in \mathcal{O}(\varepsilon^{\nu\wedge 2})$. Recall the notation in Definition 1.10. With C_ℓ a Lipschitz constant for ℓ, and some fixed $\bar z \in \mathcal{Z}$, we have

$$\ell(x) - \mathfrak{J} = \big(\ell(x) - \ell(z)\big) + \big(\ell(z) - \ell(\bar z)\big) \ge -C_\ell |x - z| \qquad \forall z \in \mathcal{S}, \forall x \in \mathbb{R}^d,$$

since $\ell(z) - \ell(\bar z) \ge 0$ for all $z \in \mathcal{S}$. Therefore, we obtain

$$\ell(x) - \mathfrak{J} \ge -C_\ell \operatorname{dist}(x,\mathcal{S}) \qquad \forall x \in \mathbb{R}^d, \tag{4.6}$$

and using the Cauchy–Schwarz inequality, and the assumption that $\ell(x) > \mathfrak{J}$ on G^c, we deduce from (4.6) and Theorem 2.2(iv) that

$$\int_{\mathbb{R}^d} \ell \,\mathrm{d}\eta_*^\varepsilon - \mathfrak{J} \ge \int_G \big(\ell(x) - \mathfrak{J}\big)\,\mathrm{d}\eta_*^\varepsilon \ge -\frac{C_\ell}{\sqrt{C_0}}\tilde\zeta^\varepsilon. \tag{4.7}$$

Thus by (4.7) and the nonnegativity of $\mathcal{G}^\varepsilon$ we have

$$-\frac{C_\ell}{\sqrt{C_0}}\tilde\zeta^\varepsilon \le \beta_*^\varepsilon - \mathfrak{J}. \tag{4.8}$$

By (4.7)–(4.8), we obtain

$$\begin{aligned} \mathcal{G}^\varepsilon_* &\le \beta^\varepsilon_* - \int_{\mathbb{R}^d} \ell \,\mathrm{d}\eta^\varepsilon_* \\ &\le \beta^\varepsilon_* - \mathfrak{J} + \mathfrak{J} - \int_{\mathbb{R}^d} \ell \,\mathrm{d}\eta^\varepsilon_* \\ &\le \beta^\varepsilon_* - \mathfrak{J} + \frac{C_\ell}{\sqrt{C_0}}\tilde{\zeta}^\varepsilon. \end{aligned} \tag{4.9}$$

An application of Young's inequality to (4.2), results in

$$\frac{C_0}{2}(\tilde{\zeta}^\varepsilon)^2 - \frac{1}{C_0}\varepsilon^2\mathcal{G}^\varepsilon_* - \varepsilon^{2\nu}\tilde{\xi}^\varepsilon \le 0,$$

and thus

$$\tilde{\zeta}^\varepsilon \le \frac{\sqrt{2}}{C_0}\varepsilon\sqrt{\mathcal{G}^\varepsilon_*} + \frac{\sqrt{2}}{\sqrt{C_0}}\varepsilon^\nu\sqrt{|\tilde{\xi}^\varepsilon|}. \tag{4.10}$$

Combining (4.9)–(4.10), and using again Young's inequality in the form

$$\frac{C_\ell}{\sqrt{C_0}}\frac{\sqrt{2}}{C_0}\varepsilon\sqrt{\mathcal{G}^\varepsilon_*} \le \frac{C_\ell^2}{C_0^3}\varepsilon^2 + \frac{1}{2}\mathcal{G}^\varepsilon_*,$$

and rearranging terms, we have

$$\frac{1}{2}\mathcal{G}^\varepsilon_* \le \beta^\varepsilon_* - \mathfrak{J} + \frac{C_\ell^2}{C_0^3}\varepsilon^2 + \frac{\sqrt{2}C_\ell}{C_0}\varepsilon^\nu\sqrt{|\tilde{\xi}^\varepsilon|}. \tag{4.11}$$

By Lemma 3.3 and (4.11), we obtain $\mathcal{G}^\varepsilon_* \in \mathcal{O}(\varepsilon^{\nu\wedge 2})$ if $\mathfrak{J} = \mathfrak{J}_{\mathrm{s}}$ for $\nu > 1$, or if $\mathfrak{J}_{\mathrm{c}} = \mathfrak{J}_{\mathrm{s}}$ and $\nu = 1$. We also obtain $\mathcal{G}^\varepsilon_* \in \mathcal{O}(\varepsilon^{2\wedge(2\nu-2)})$ if $\mathfrak{J} < \mathfrak{J}_s$ and $\nu > 1$. Thus (4.4) holds.

If $\mathfrak{J} < \mathfrak{J}_{\mathrm{s}}$ and $\nu > 1$, then $\mathcal{Z} \subset \mathcal{S}\setminus\mathcal{S}_{\mathrm{s}}$ and $\mathfrak{S} \subset \mathcal{Z}$ by Lemma 3.3. Fix some $z \in \mathfrak{S}$. Then $\liminf_{\varepsilon\searrow 0} \eta^\varepsilon_*(B_r(z)) > 0$ for any $r > 0$. Also, $\Delta\mathcal{V}(z) < 0$ by Theorem 2.2. Therefore, (3.42) holds, with "lim inf" replacing the "lim sup". Expanding $\tilde{\xi}^\varepsilon_1$ as in (3.41), and arguing as in Lemma 3.6 it follows that (3.42) with "lim inf" also holds for $\tilde{\xi}^\varepsilon$. In fact, it easily follows that for some constant κ_1, we have

$$\liminf_{\varepsilon\searrow 0}(-\tilde{\xi}^\varepsilon) \ge \min_{z\in\mathcal{Z}} \kappa_1\left(-\frac{1}{2}\Delta\mathcal{V}(z)\right). \tag{4.12}$$

The discriminant of the quadratic polynomial in (4.2) is nonnegative and this implies that

$$\varepsilon^2\mathcal{G}^\varepsilon_* \ge -2C_0\varepsilon^{2\nu}\tilde{\xi}^\varepsilon, \tag{4.13}$$

in direct analogy with (3.43). Thus, (4.5) follows by (4.12) and (4.13). This completes the proof of part (a).

Since $\tilde{\zeta}^\varepsilon \in \mathcal{O}(\varepsilon^{\nu\wedge 2})$, we obtain $\beta_*^\varepsilon - \mathfrak{J} \geq \mathcal{O}(\varepsilon^{\nu\wedge 2})$ by (4.8). This proves part (b), and completes the proof. $\square$

We define the following scaled quantities.

DEFINITION 4.3. For $z \in \mathcal{S}$, and V^ε as in Theorem 1.4, we define

$$\widehat{V}_z^\varepsilon(x) := V^\varepsilon\big(\varepsilon^\nu x + z\big), \qquad x \in \mathbb{R}^d$$

and

$$\widetilde{V}^\varepsilon := \varepsilon^2 V^\varepsilon, \qquad \check{V}_z^\varepsilon := \varepsilon^{2(1-\nu)}\widehat{V}_z^\varepsilon.$$

We also define the "scaled" vector field and penalty by

$$\widehat{m}_z^\varepsilon(x) := \frac{m(\varepsilon^\nu x + z)}{\varepsilon^\nu} \quad \text{and} \quad \widehat{\ell}_z^\varepsilon(x) := \ell\big(\varepsilon^\nu x + z\big),$$

respectively.

The next lemma provides estimates for the growth of $\nabla\widehat{V}_z^\varepsilon$ and $\nabla\widetilde{V}^\varepsilon$.

LEMMA 4.4. *Assume $\nu \in (0,2]$, and let $\widehat{V}_z^\varepsilon$, $\widetilde{V}^\varepsilon$, and $\check{V}_z^\varepsilon$, be as in Definition 4.3. We have the following.*

(a) *Under the restriction that $z \in \mathcal{Z}$ when $\nu \in (1,2]$, there exists a constant $\check{c}_0$ such that*

$$\big|\nabla\check{V}_z^\varepsilon(x)\big| \leq \check{c}_0\big(1+|x|\big) \qquad \forall\varepsilon \in (0,1), \forall x \in \mathbb{R}^d. \tag{4.14}$$

(b) *The bound in (4.14) also holds for $\widetilde{V}^\varepsilon$ for all $\nu \in (0,2]$, with no restrictions on z.*

PROOF. By (1.16), the function $\check{V}_z^\varepsilon$ satisfies

$$\frac{1}{2}\Delta\check{V}_z^\varepsilon(x) + \big\langle\widehat{m}_z^\varepsilon(x), \nabla\check{V}_z^\varepsilon(x)\big\rangle - \frac{1}{2}\big|\nabla\check{V}_z^\varepsilon(x)\big|^2 = \varepsilon^{2(1-\nu)}\big(\beta_*^\varepsilon - \widehat{\ell}_z^\varepsilon(x)\big). \tag{4.15}$$

Since ℓ is Lipschitz, the gradient of the map $x \mapsto \varepsilon^{2(1-\nu)}(\widehat{\ell}_z^\varepsilon(x) - \ell(z))$ is bounded in $\mathbb{R}^d$, uniformly in $\varepsilon \in (0,1)$, and $\nu \in (0,2]$. Similarly, $|\widehat{m}_z^\varepsilon(x)|$, $\|D\widehat{m}_z^\varepsilon(x)\|$, and $\|D^2\widehat{m}_z^\varepsilon(x)\|$, are bounded in $\mathbb{R}^d$, uniformly in $\varepsilon \in (0,1)$ and $\nu \in (0,2]$. By Theorem 1.11(i), which is established in Corollary 4.2, the constants $\varepsilon^{2(1-\nu)}(\beta_*^\varepsilon - \ell(z))$ are bounded uniformly in $\varepsilon \in (0,1)$, and $\nu \in (1,2]$ for $z \in \mathcal{Z}$. Applying Metafune, Pallara and Rhandi (2005), Lemma 5.1 to (4.15) it follows that $\check{V}_z^\varepsilon$ satisfies (4.14) if $\nu \in (1,2]$ and $z \in \mathcal{Z}$. On the other hand, if $\nu \in (0,1]$, then the gradient of the right-hand side of (4.15) is bounded in $\mathbb{R}^d$, uniformly in $\varepsilon \in (0,1)$, and the restriction $z \in \mathcal{Z}$ is not needed. This completes the proof of part (a).

We next show that (4.14) holds for $\widetilde{V}^\varepsilon$. Fix an arbitrary $z \in \mathcal{Z}$. We have

$$\begin{aligned}\nabla_x V^\varepsilon(x+z) &= \varepsilon^{-\nu} \nabla_y \widehat{V}_z^\varepsilon(y)|_{y=\varepsilon^{-\nu}x} \\ &\le \frac{\varepsilon^{-\nu}}{\varepsilon^{2(1-\nu)}} \breve{c}_0\big(1+|\varepsilon^{-\nu}x|\big) = \frac{\breve{c}_0}{\varepsilon^2}\big(\varepsilon^\nu + |x|\big),\end{aligned}$$

where in the inequality we use the identity $\widehat{V}_z^\varepsilon = \varepsilon^{2(\nu-1)} \breve{V}_z^\varepsilon$ and (4.14). Since $\widetilde{V}^\varepsilon = \varepsilon^2 V^\varepsilon$, this establishes part (b) and completes the proof. □

We continue with a version of Lemma 4.1 for unbounded domains.

PROPOSITION 4.5. *Let $\nu \in (0,2]$. Then for any $k \in \mathbb{N}$ and $r > 0$, there exist constants and $\hat{\kappa}_1 = \hat{\kappa}_1(k)$ and $\hat{\kappa}_2 = \hat{\kappa}_1(k)$ such that with $\hat{r}(\varepsilon) := \hat{\kappa}_2 \varepsilon^{\nu \wedge 1}$ we have*

$$\int_{B^c_{\hat{r}(\varepsilon)}(\mathcal{S})} \big(\operatorname{dist}(x,\mathcal{S})\big)^{2k} \eta_*^\varepsilon(\mathrm{d}x) \le \hat{\kappa}_1 \varepsilon^{2(\nu\wedge 1)} \qquad \forall \varepsilon \in (0,1).$$

PROOF. Let $\widetilde{V}^\varepsilon := \varepsilon^2 V^\varepsilon$. Since $V^\varepsilon(0) = 0$, the function $\widetilde{V}^\varepsilon = \varepsilon^2 V^\varepsilon$ is locally bounded, uniformly in $\varepsilon > 0$, by Lemma 4.4. Applying the operator

$$\mathcal{L}_*^\varepsilon := \frac{\varepsilon^{2\nu}}{2}\Delta + \langle m - \varepsilon^2 \nabla V^\varepsilon, \nabla \rangle$$

to the function $\mathcal{V}^{2k} \mathrm{e}^{\widetilde{V}^\varepsilon}$ and using the identity $\mathcal{L}_*^\varepsilon[\widetilde{V}^\varepsilon] = \varepsilon^2(\beta_*^\varepsilon - \ell) - \frac{1}{2}|\nabla \widetilde{V}^\varepsilon|^2$, and rearranging terms, we obtain

$$\begin{aligned}&\mathcal{L}_*^\varepsilon[\mathcal{V}^{2k}\mathrm{e}^{\widetilde{V}^\varepsilon}] \\ &\quad= \mathcal{V}^{2k}\mathcal{L}_*^\varepsilon[\mathrm{e}^{\widetilde{V}^\varepsilon}] + \mathrm{e}^{\widetilde{V}^\varepsilon}\mathcal{L}_*^\varepsilon[\mathcal{V}^{2k}] + 2k\varepsilon^{2\nu}\mathcal{V}^{(2k-1)}\mathrm{e}^{\widetilde{V}^\varepsilon}\langle \nabla \widetilde{V}^\varepsilon, \nabla \mathcal{V}\rangle \\ &\quad= \mathcal{V}^{2k}\mathrm{e}^{\widetilde{V}^\varepsilon}\left(\varepsilon^2(\beta_*^\varepsilon - \ell) + \frac{\varepsilon^{2\nu}-1}{2}|\nabla \widetilde{V}^\varepsilon|^2\right) + \mathrm{e}^{\widetilde{V}^\varepsilon}\big(k\varepsilon^{2\nu}\mathcal{V}^{2k-1}\Delta\mathcal{V} \\ &\qquad + k(2k-1)\varepsilon^{2\nu}\mathcal{V}^{2k-2}|\nabla\mathcal{V}|^2 + 2k\mathcal{V}^{2k-1}\langle m - \varepsilon^2\nabla V^\varepsilon, \nabla\mathcal{V}\rangle\big) \\ &\qquad + 2k\varepsilon^{2\nu}\mathcal{V}^{(2k-1)}\mathrm{e}^{\widetilde{V}^\varepsilon}\langle \nabla\widetilde{V}^\varepsilon, \nabla\mathcal{V}\rangle \qquad (4.16) \\ &\quad= \mathcal{V}^{2k}\mathrm{e}^{\widetilde{V}^\varepsilon}\left[\varepsilon^2(\beta_*^\varepsilon - \ell) + k\varepsilon^{2\nu}\frac{\Delta\mathcal{V}}{\mathcal{V}} - \frac{1-\varepsilon^{2\nu}}{2}|\nabla\widetilde{V}^\varepsilon|^2\right. \\ &\qquad \left. - \frac{2k(1-\varepsilon^{2\nu})}{\mathcal{V}}\langle\nabla\widetilde{V}^\varepsilon, \nabla\mathcal{V}\rangle + 2k\frac{\langle m, \nabla\mathcal{V}\rangle}{\mathcal{V}} + k(2k-1)\varepsilon^{2\nu}\frac{|\nabla\mathcal{V}|^2}{\mathcal{V}^2}\right] \\ &\quad= \mathcal{V}^{2k}\mathrm{e}^{\widetilde{V}^\varepsilon}\left[\varepsilon^2(\beta_*^\varepsilon - \ell) + k\varepsilon^{2\nu}\frac{\Delta\mathcal{V}}{\mathcal{V}} - \frac{1-\varepsilon^{2\nu}}{2}\left(\nabla\widetilde{V}^\varepsilon + 2k\frac{\nabla\mathcal{V}}{\mathcal{V}}\right)^2\right. \\ &\qquad \left. + 2k\frac{\langle m, \nabla\mathcal{V}\rangle}{\mathcal{V}} + k\big(2k - \varepsilon^{2\nu}\big)\frac{|\nabla\mathcal{V}|^2}{\mathcal{V}^2}\right].\end{aligned}$$

By (2.7), and since $\bar{\mathcal{V}}$ has strict quadratic growth and $\nabla\bar{\mathcal{V}}$ is Lipschitz by Hypothesis 1.1, and $\mathcal{V}$ agrees with $\bar{\mathcal{V}}$ outside a compact set, it follows that $\frac{|\nabla\mathcal{V}|^2}{\mathcal{V}}$ is bounded on $\mathbb{R}^d$. Therefore, in view of the bounds in (2.1) and (2.7), we can add a positive constant to $\mathcal{V}$ so that

$$2\frac{\langle m,\nabla\mathcal{V}\rangle}{\mathcal{V}}+\bigl(2k-\varepsilon^{2\nu}\bigr)\frac{|\nabla\mathcal{V}|^2}{\mathcal{V}^2}\le\frac{\langle m,\nabla\mathcal{V}\rangle}{\mathcal{V}}\qquad\text{on }\mathbb{R}^d,\ \forall\varepsilon>0. \tag{4.17}$$

This constant may be selected so that $\mathcal{V}\ge 1$ on $\mathbb{R}^d$. Define

$$G_0^\varepsilon:=\varepsilon^{2-2\wedge 2\nu}\bigl(\beta_*^\varepsilon-\ell\bigr)-\frac{1-\varepsilon^{2\nu}}{2\varepsilon^{2\wedge 2\nu}}\left|\nabla\widetilde{V}^\varepsilon+2k\frac{\nabla\mathcal{V}}{\mathcal{V}}\right|^2.$$

Since ℓ is inf-compact, there exists $r_0>0$ such that $G_0^\varepsilon\le 0$ on $B_{r_0}^c$. We may choose r_0 large enough so that $\mathcal{S}\subset B_{r_0}$. Let κ_0 be a bound of $\beta_*^\varepsilon-\ell$ on B_{r_0}. Using this bound and (4.16)–(4.17), we obtain

$$\begin{aligned}&\frac{1}{\varepsilon^{2\wedge 2\nu}}\mathcal{L}_*^\varepsilon\bigl[\mathcal{V}^{2k}\mathrm{e}^{\widetilde{V}^\varepsilon}\bigr](x)\\&\qquad\le\mathcal{V}^{2k}(x)\mathrm{e}^{\widetilde{V}^\varepsilon(x)}\left[\kappa_0\mathbb{1}_{B_{r_0}}(x)+\frac{k}{\varepsilon^{2\wedge 2\nu}}\frac{\varepsilon^{2\nu}\Delta\mathcal{V}(x)+\langle m(x),\nabla\mathcal{V}(x)\rangle}{\mathcal{V}(x)}\right]\end{aligned} \tag{4.18}$$

for all $x\in\mathbb{R}^d$, and all $\varepsilon\in(0,1)$. By (2.1) we have

$$\varepsilon^{2\nu}\Delta\mathcal{V}(x)+\bigl\langle m(x),\nabla\mathcal{V}(x)\bigr\rangle\le\frac{1}{2}\bigl\langle m(x),\nabla\mathcal{V}(x)\bigr\rangle \tag{4.19}$$

for all $x\in\mathbb{R}^d$ such that $\operatorname{dist}(x,\mathcal{S})\ge\kappa_1\varepsilon^\nu$, with $\kappa_1:=\sqrt{2C_0^{-1}\|\Delta\mathcal{V}\|_\infty}$. Using (2.1) once more, if we define $\kappa_2:=(4k^{-1}C_0^{-1}\kappa_0\sup_{B_{r_0}}\mathcal{V})^{\frac{1}{2}}$, then we have

$$\varepsilon^{2\wedge 2\nu}\kappa_0+\frac{k\langle m(x),\nabla\mathcal{V}(x)\rangle}{4\mathcal{V}(x)}\le 0 \tag{4.20}$$

in $\{x\in B_{r_0}\colon\ \operatorname{dist}(x,\mathcal{S})\ge\kappa_2\varepsilon^{1\wedge\nu}\}$. Combining (4.18), (4.19), and (4.20), we obtain

$$\frac{1}{\varepsilon^{2\wedge 2\nu}}\mathcal{L}_*^\varepsilon\bigl[\mathcal{V}^{2k}\mathrm{e}^{\widetilde{V}^\varepsilon}\bigr](x)\le\frac{k}{4\varepsilon^{2\wedge 2\nu}}\mathcal{V}^{2k-1}(x)\mathrm{e}^{\widetilde{V}^\varepsilon(x)}\bigl\langle m(x),\nabla\mathcal{V}(x)\bigr\rangle \tag{4.21}$$

for all $x\in\mathbb{R}^d$ such that $\operatorname{dist}(x,\mathcal{S})\ge\hat{r}(\varepsilon):=(\kappa_1\vee\kappa_2)\varepsilon^{\nu\wedge 1}$. Let κ_3 be a bound of the right-hand side of (4.18) on $B_{\hat{r}(\varepsilon)}(\mathcal{S})$. This bound does not depend on ε, since $\widetilde{V}^\varepsilon$ is locally bounded, uniformly in $\varepsilon\in(0,1)$. Then, by (4.18) and (4.21) we obtain

$$\begin{aligned}&\frac{1}{\varepsilon^{2\wedge 2\nu}}\mathcal{L}_*^\varepsilon\bigl[\mathcal{V}^{2k}\mathrm{e}^{\widetilde{V}^\varepsilon}\bigr](x)\\&\qquad\le\kappa_3+\frac{k}{4\varepsilon^{2\wedge 2\nu}}\bigl\langle m(x),\nabla\mathcal{V}(x)\bigr\rangle\mathcal{V}^{2k-1}(x)\mathrm{e}^{\widetilde{V}^\varepsilon(x)}\mathbb{1}_{B^c_{\hat{r}(\varepsilon)}(\mathcal{S})}(x)\end{aligned} \tag{4.22}$$

for all $x\in\mathbb{R}^d$, and $\varepsilon\in(0,1)$.

By the strong maximum principle, V^ε attains its infimum in $\mathbb{R}^d$ in the set $\{x \in \mathbb{R}^d : \ell(x) \le \beta^\varepsilon_*\}$. Therefore, $\widetilde{V}^\varepsilon$ is bounded below in $\mathbb{R}^d$, uniformly in ε, by Lemma 4.4. Thus, from (4.22) we obtain

$$\int_{B^c_{\hat{r}(\varepsilon)}(\mathcal{S})} \frac{|\langle m(x), \nabla \mathcal{V}(x)\rangle|}{\varepsilon^{2\wedge 2\nu}} \mathcal{V}^{2k-1}(x)\eta^\varepsilon_*(\mathrm{d}x) \le \frac{4\kappa_3}{k(\inf_{\mathbb{R}^d} \mathrm{e}^{\widetilde{V}^\varepsilon})} \tag{4.23}$$

for all $\varepsilon \in (0,1)$. By the strict quadratic growth of $\mathcal{V}$ mentioned earlier, together with (2.7) and (4.23), there exists a constant κ_4, such that

$$\int_{B^c_{\hat{r}(\varepsilon)}(\mathcal{S})} \frac{1}{\varepsilon^{2\wedge 2\nu}} \big(\operatorname{dist}(x,\mathcal{S})\big)^{4k-1} \eta^\varepsilon_*(\mathrm{d}x) \le \kappa_4 \qquad \forall \varepsilon \in (0,1).$$

This, together with Lemma 4.1 finishes the proof. □

COROLLARY 4.6. *Let D be any open set such that $\mathcal{S}_{\mathrm{s}} \subset D$. The following hold.*

(a) *If $\mathfrak{J} = \mathfrak{J}_{\mathrm{s}}$, then $\eta^\varepsilon_*(D^c) \in \mathcal{O}(\varepsilon^{2-\nu})$ for all $\nu \in (1,2)$.*
(b) *If $\nu \in (0,1)$, then*

$$\mathcal{G}^\varepsilon_* \in \mathcal{O}\big(\varepsilon^\nu\big), \qquad \beta^\varepsilon_* - \mathfrak{J} \ge \mathcal{O}\big(\varepsilon^\nu\big) \quad \text{and} \quad \eta^\varepsilon_*\big(D^c\big) \in \mathcal{O}\big(\varepsilon^{2\nu\wedge(2-\nu)}\big). \tag{4.24}$$

PROOF. Since $2-\nu < 2(\nu\wedge 1)$ for $\nu \in [1,2)$, then, in view of Proposition 4.5, it suffices to prove that $\eta^\varepsilon_*(B_r(z)) \in \mathcal{O}(\varepsilon^{2-\nu})$ for a bounded open neighborhood $B_r(z)$ of $z \in \mathcal{S} \setminus \mathcal{S}_{\mathrm{s}}$. Let φ be as in the proof of Lemma 3.6, and ξ^ε_i, $i = 1,2$, as defined in (3.38). By Proposition 4.5, we have

$$\xi^\varepsilon_2 \in \mathcal{O}\big(\varepsilon^{2(\nu\wedge 1)}\big) \quad \text{and} \quad \int_{B^c_r(\mathcal{S})} \varphi'(\mathcal{V})\Delta\mathcal{V}\,\mathrm{d}\eta^\varepsilon_* \in \mathcal{O}\big(\varepsilon^{2(\nu\wedge 1)}\big).$$

Thus

$$\xi^\varepsilon \le \frac{1}{2}\Delta\mathcal{V}(z)\eta^\varepsilon_*\big(B_r(z)\big) + \mathcal{O}\big(\varepsilon^{2(\nu\wedge 1)}\big), \qquad z \in \mathcal{S}\setminus\mathcal{S}_{\mathrm{s}}, \tag{4.25}$$

by (3.36) and (3.41). In addition, we have $\mathcal{G}^\varepsilon_* \in \mathcal{O}(\varepsilon^{\nu\wedge 2})$ by Corollary 4.2(a), and $-C_0\xi^\varepsilon \le \frac{1}{2}\varepsilon^{2-2\nu}\mathcal{G}^\varepsilon_*$ by (3.43). We combine these with (4.25) for $\nu \in (1,2)$ to obtain

$$-C_0\Delta\mathcal{V}(z)\eta^\varepsilon_*\big(B_r(z)\big) + \mathcal{O}\big(\varepsilon^2\big) \le \varepsilon^{2-2\nu}\mathcal{G}^\varepsilon_* \in \mathcal{O}\big(\varepsilon^{2-\nu}\big).$$

Thus, $\eta^\varepsilon_*(B_r(z)) \in \mathcal{O}(\varepsilon^{2-\nu})$ for $\nu \in (1,2)$. This completes the proof of part (a).

The proof of part (b) is divided in two steps.

Step 1. Suppose $\mathfrak{J} = \mathfrak{J}_{\mathrm{s}}$. Then (4.8)–(4.11) hold with $\mathfrak{J}$ replaced by $\mathfrak{J}_{\mathrm{s}}$. By Lemma 3.5(ii) we have $\beta^\varepsilon_* - \mathfrak{J}_{\mathrm{s}} \le \mathcal{O}(\varepsilon^{\nu\vee(4\nu-2)})$. It follows that $\mathcal{G}^\varepsilon_* \in \mathcal{O}(\varepsilon^\nu)$ by (4.11), and thus $\tilde{\zeta}^\varepsilon \in \mathcal{O}(\varepsilon^\nu)$ by (4.10). Hence, $\beta^\varepsilon_* - \mathfrak{J}_{\mathrm{s}} \ge \mathcal{O}(\varepsilon^\nu)$ by (4.8).

By (4.25) and (3.43), and since $\nu \in (0,1)$, we obtain

$$-C_0\Delta\mathcal{V}(z)\eta^\varepsilon_*\big(B_r(z)\big) + \mathcal{O}\big(\varepsilon^{2\nu}\big) \le \varepsilon^{2-2\nu}\mathcal{G}^\varepsilon_*.$$

As already shown, $\mathcal{G}^\varepsilon_* \in \mathcal{O}(\varepsilon^\nu)$. Combining these estimates with Proposition 4.5, we obtain $\eta^\varepsilon_*(D^c) \in \mathcal{O}(\varepsilon^{2\nu\wedge(2-\nu)})$.

Step 2. Suppose $\mathfrak{J} < \mathfrak{J}_{\rm s}$. By Theorem 2.2(ii), we may construct $\mathcal{V}$ such that $\mathcal{V}(z) > 5\max_{\mathcal{S}_{\rm s}} \mathcal{V}$ for all $z \in \mathcal{S} \setminus \mathcal{S}_{\rm s}$. Let

$$G := \Big\{ x \in \mathbb{R}^d \colon \mathcal{V}(x) < 2 \max_{\mathcal{S}_{\rm s}} \mathcal{V} \Big\},$$

and $\tilde{\varphi}$ be as in the proof of Lemma 4.1, with $\delta = 2\max_{\mathcal{S}_{\rm s}} \mathcal{V}$. We have

$$\mathfrak{J}_{\rm s} - \ell(x) \le \ell(z) - \ell(x) \le C_\ell |x - z| \qquad \forall z \in \mathcal{S}_{\rm s}, \text{ and } x \in \mathbb{R}^d.$$

Thus

$$\ell(x) - \mathfrak{J}_{\rm s} \ge \max_{z \in \mathcal{S}_{\rm s}} \{ -C_\ell |x - z| \} = -C_\ell \operatorname{dist}(x, \mathcal{S}_{\rm s}) \qquad \forall x \in \mathbb{R}^d.$$

By Proposition 4.5, for some positive constants r and κ_1 we obtain

$$\int_{G^c} \big(\ell(x) - \mathfrak{J}_{\rm s}\big)\, \mathrm{d}\eta^\varepsilon_* \ge -\kappa_1 \sum_{z \in \mathcal{S} \setminus \mathcal{S}_{\rm s}} \eta^\varepsilon_*\big(B_r(z)\big) + \mathcal{O}\big(\varepsilon^{2\nu}\big).$$

Therefore, splitting the integral over G and G^c, we obtain as in (4.7) that

$$\int_{\mathbb{R}^d} \ell \, \mathrm{d}\eta^\varepsilon_* - \mathfrak{J}_{\rm s} \ge -\kappa_1 \sum_{z \in \mathcal{S} \setminus \mathcal{S}_{\rm s}} \eta^\varepsilon_*\big(B_r(z)\big) + \mathcal{O}\big(\varepsilon^{2\nu}\big) - \frac{C_\ell}{\sqrt{C_0}} \tilde{\zeta}^\varepsilon,$$

and since $\tilde{\zeta}^\varepsilon \in \mathcal{O}(\varepsilon^\nu)$, following the steps in (4.8)–(4.11) we have

$$-\kappa_1 \sum_{z \in \mathcal{S} \setminus \mathcal{S}_{\rm s}} \eta^\varepsilon_*\big(B_r(z)\big) - \mathcal{O}\big(\varepsilon^{2\nu}\big) - \frac{C_\ell}{\sqrt{C_0}} \tilde{\zeta}^\varepsilon \le \beta^\varepsilon_* - \mathfrak{J}_{\rm s}, \tag{4.26}$$

and

$$\begin{aligned} \frac{1}{2} \mathcal{G}^\varepsilon_* \le \beta^\varepsilon_* - \mathfrak{J}_{\rm s} + \frac{C_\ell}{C_0^3} \varepsilon^2 + \frac{\sqrt{2} C_\ell}{C_0} \varepsilon^\nu \sqrt{|\tilde{\xi}^\varepsilon|} \\ + \kappa_1 \sum_{z \in \mathcal{S} \setminus \mathcal{S}_{\rm s}} \eta^\varepsilon_*\big(B_r(z)\big) + \mathcal{O}\big(\varepsilon^{2\nu}\big). \end{aligned} \tag{4.27}$$

Consider now a point $z \in \mathcal{S} \setminus \mathcal{S}_{\rm s}$, and let φ be the function constructed in Lemma 3.6 relative to z. Recall the definitions in (3.38). It is a direct consequence of (3.43) and (4.25) that

$$\sum_{z \in \mathcal{S} \setminus \mathcal{S}_{\rm s}} \eta^\varepsilon_*\big(B_r(z)\big) \le \kappa_2 \big(\varepsilon^{2-2\nu} \mathcal{G}^\varepsilon_* + \varepsilon^{2\nu}\big) \tag{4.28}$$

for some positive constant κ_2. Since $\beta^\varepsilon_* - \mathfrak{J}_{\rm s} \le \mathcal{O}(\varepsilon^{\nu \vee (4\nu-2)})$ by Lemma 3.6, and $\nu < 1$, combining (4.27) and (4.28) we obtain $\mathcal{G}^\varepsilon_* \in \mathcal{O}(\varepsilon^\nu)$. Therefore, by (4.28), we obtain $\eta^\varepsilon_*(B_r(z)) \in \mathcal{O}(\varepsilon^{2\nu\wedge(2-\nu)})$ for all $z \in \mathcal{S} \setminus \mathcal{S}_{\rm s}$. In turn, $\beta^\varepsilon_* - \mathfrak{J} \ge \mathcal{O}(\varepsilon^\nu)$ by (4.26). This completes the proof. □

REMARK 4.7. If $\nu = 1$ and $\mathfrak{J}_c = \mathfrak{J}_s$, then following the argument in Step 2 of the proof of Corollary 4.6 we obtain the same estimates as in (4.24). In this case, we do not estimate $\mathcal{G}^\varepsilon_*$ from (4.27), but rather use Corollary 4.2(a) which asserts that $\mathcal{G}^\varepsilon_* \in \mathcal{O}(\varepsilon)$. Thus $\eta^\varepsilon_*(B_r(z)) \in \mathcal{O}(\varepsilon)$ by (4.28), which, in turn, implies that $\beta^\varepsilon_* - \mathfrak{J}_s \geq \mathcal{O}(\varepsilon)$ by (4.26).

5. Convergence of the scaled optimal stationary distributions. We need the following definition.

DEFINITION 5.1. For the rest of the paper $\{\mathcal{B}_z : z \in \mathcal{S}\}$ is some collection of nonempty, disjoint balls, with each $\mathcal{B}_z$ centered around z, and we define $\mathcal{B}_\mathcal{S} := \bigcup_{z\in\mathcal{S}} \mathcal{B}_z$.

Recall $\widehat{V}^\varepsilon_z$ from Definition 4.3. For $z \in \mathcal{S}$, we define the "scaled" density $\hat{\varrho}^\varepsilon_z(x) := \varepsilon^{\nu d}\varrho^\varepsilon_*(\varepsilon^\nu x + z)$, and denote by $\hat{\eta}^\varepsilon_z$ the corresponding probability measure in $\mathbb{R}^d$. We also define the "normalized" probability density $\mathring{\varrho}^\varepsilon_z$ supported on $\eta^\varepsilon_*(\mathcal{B}_z)$ by

$$\mathring{\varrho}^\varepsilon_z(x) := \begin{cases} \dfrac{\hat{\varrho}^\varepsilon_z(x)}{\eta^\varepsilon_*(\mathcal{B}_z)} & \text{if } \varepsilon^\nu x + z \in \mathcal{B}_z, \\ 0 & \text{otherwise,} \end{cases}$$

and let $\mathring{\eta}^\varepsilon_z(\mathrm{d}x) = \mathring{\varrho}^\varepsilon_z(x)\,\mathrm{d}x$.

Section 5.1 which follows concerns the critical regime. The subcritical and supercritical regimes are treated in Section 5.2.

5.1. *Convergence to a Gaussian in the critical regime.* Recall the notation in Definitions 1.9 and 1.10, and the scaled quantities in Definition 4.3. We start with the following lemma.

LEMMA 5.2. *Assume $\nu = 1$. Fix any $z \in \mathcal{S}$. Then every sequence $\varepsilon_n \searrow 0$ has a subsequence along which $\widehat{V}^\varepsilon_z(x) - \widehat{V}^\varepsilon_z(z)$ converges to some $\bar{V}_z \in \mathcal{C}^2(\mathbb{R}^d)$ uniformly on compact subsets of $\mathbb{R}^d$, and β^ε_* converges to some constant $\bar{\beta}_*$, and these satisfy*

$$\frac{1}{2}\Delta\bar{V}_z(x) + \langle M_z x, \nabla\bar{V}_z(x)\rangle - \frac{1}{2}\big|\nabla\bar{V}_z(x)\big|^2 = \bar{\beta}_* - \ell(z). \tag{5.1}$$

Moreover, for some constant $\hat{c}_0$ we have

$$\big|\nabla\bar{V}_z(x)\big| \leq \hat{c}_0\big(1 + |x|\big) \qquad \forall\varepsilon \in (0,1), \forall x \in \mathbb{R}^d \tag{5.2}$$

and

$$\bar{\beta}_* \leq \Lambda^+(M_z) + \ell(z). \tag{5.3}$$

PROOF. If $\nu = 1$, then by (4.15) we obtain

$$\frac{1}{2}\Delta \widehat{V}_z^\varepsilon + \langle \widehat{m}_z^\varepsilon, \nabla \widehat{V}_z^\varepsilon \rangle - \frac{1}{2}|\nabla \widehat{V}_z^\varepsilon|^2 + \widehat{\ell}_z^\varepsilon = \beta_*^\varepsilon. \tag{5.4}$$

By applying Metafune, Pallara and Rhandi (2005), Lemma 5.1 to (5.4) and using the assumptions on the growth of m and ℓ, it follows that there exists a constant $\hat{c}_0$ such that

$$\bigl|\nabla \widehat{V}_z^\varepsilon(x)\bigr| \le \hat{c}_0\bigl(1 + |x|\bigr) \qquad \forall \varepsilon \in (0,1), \forall x \in \mathbb{R}^d. \tag{5.5}$$

It follows by (5.4) and the bound in (5.5) that $\widehat{V}_z^\varepsilon$ is locally bounded in $\mathcal{C}^{2,\alpha}(\mathbb{R}^d)$ for any $\alpha \in (0,1)$. It is also clear that $\widehat{m}_z^\varepsilon(x) \to M_z x$ and $\widehat{\ell}_z^\varepsilon(x) \to \ell(z)$, as $\varepsilon \searrow 0$, uniformly over compact sets. Thus, taking limits in (5.4) along some sequence $\varepsilon_n \searrow 0$, we obtain a function $\bar{V}_z \in \mathcal{C}^2(\mathbb{R}^d)$ and a constant $\bar{\beta}_*$ which satisfy (5.1). The bound in (5.2) follows by (5.5), while the bound in (5.3) follows by applying Theorem 1.19(c) to (5.1), with $\bar{\beta} = \bar{\beta}_* - \ell(z)$. $\square$

We fix some notation. The function $\bar{V}_z$ for $z \in \mathcal{S}$ denotes the limit obtained in Lemma 5.2. The associated "diffusion limit", takes the form

$$\mathrm{d}\bar{X}_t = \bigl(M_z \bar{X}_t - \nabla \bar{V}_z(\bar{X}_t)\bigr)\,\mathrm{d}t + \mathrm{d}\bar{W}_t, \tag{5.6}$$

and its extended generator is denoted by

$$\bar{\mathcal{L}}_z := \frac{1}{2}\Delta + \bigl\langle M_z x - \nabla \bar{V}_z(x), \nabla \bigr\rangle. \tag{5.7}$$

Since (5.3) holds for all $z \in \mathcal{S}$, then we must have $\bar{\beta}_* \le \mathfrak{J}_{\mathrm{c}}$, and Lemma 5.2 provides an alternate proof of the upper bound $\limsup_{\varepsilon \searrow 0} \beta_*^\varepsilon \le \mathfrak{J}_{\mathrm{c}}$, which was already shown in Lemma 3.3. In the next theorem we show that if $\liminf_{\varepsilon_n \searrow 0} \eta_*^{\varepsilon_n}(\mathcal{B}_z) > 0$ over some sequence $\{\varepsilon_n\}$, then the diffusion in (5.6) is positive recurrent.

THEOREM 5.3. *Assume $\nu = 1$, and let $\{\mathcal{B}_z : z \in \mathcal{S}\}$ be as in Definition 5.1. Let $\varepsilon_n \searrow 0$ be any sequence satisfying $\liminf_{n\to\infty} \eta_*^{\varepsilon_n}(\mathcal{B}_z) = \theta_z > 0$ for some $z \in \mathcal{S}$, and let $(\bar{V}_z, \bar{\beta}_*) \in \mathcal{C}^2(\mathbb{R}^d) \times \mathbb{R}$ be any limit point of $(\widehat{V}_z^\varepsilon(x) - \widehat{V}_z^\varepsilon(z), \beta_*^\varepsilon)$ along some subsequence of $\{\varepsilon_n\}$ (see Lemma 5.2). Recall Definition 1.9. Then the following hold.*

(a) *The diffusion in (5.6) is positive recurrent with invariant probability measure $\bar{\eta}_z$, and the density $\mathring{\varrho}_z^\varepsilon$ in Definition 5.1 converges to the density $\bar{\varrho}_z$ of $\bar{\eta}_z$, uniformly on compact subsets of $\mathbb{R}^d$.*

(b) *The invariant probability measure $\bar{\eta}_z$ has finite second moments.*

(c) *It holds that $\bar{\beta}_* = \ell(z) + \Lambda^+(M_z)$.*

(d) *We have*

$$\widehat{V}_z(x) = \frac{1}{2}\langle x, \widehat{Q}_z x \rangle, \tag{5.8}$$

and that $\bar{\varrho}_z$ is the density of a Gaussian with mean 0 and covariance matrix $\widehat{\Sigma}_z$. Here, $(\widehat{Q}_z, \widehat{\Sigma}_z)$ are the pair of matrices which solve (1.17).

(e) *It holds that*

$$\liminf_{\varepsilon_n \searrow 0} \int_{\mathcal{B}_z} \Big(\ell(x) + \frac{1}{2} \big| v_*^{\varepsilon_n}(x) \big|^2 \Big) \eta_*^{\varepsilon_n}(\mathrm{d}x) \ge \theta_z \big(\ell(z) + \Lambda^+(M_z) \big).$$

PROOF. In order to show that the diffusion in (5.6) is positive recurrent, we examine the scaled diffusion

$$\mathrm{d}X_t = \big(\widehat{m}_z^\varepsilon(X_t) - \nabla \widehat{V}_z^\varepsilon(X_t) \big) \, \mathrm{d}t + \mathrm{d}W_t. \tag{5.9}$$

Recall from Definition 5.1 that $\hat{\eta}_z^\varepsilon$ and $\hat{\varrho}_z^\varepsilon$ denote the invariant probability measure of the diffusion in (5.9) and its density, respectively. Let

$$\widehat{\mathcal{L}}_z^\varepsilon := \frac{1}{2} \Delta + \langle \widehat{m}_z^\varepsilon - \nabla \widehat{V}_z^\varepsilon, \nabla \rangle$$

denote the extended generator of (5.9). It follows by Lemma 4.1 and the Markov inequality that $\eta_*^\varepsilon(\mathcal{B}_z \setminus B_{n\varepsilon}(z)) \le \frac{\hat{\kappa}_0}{n^2}$ for all $n \in \mathbb{N}$. Therefore, $\{\mathring{\eta}_z^{\varepsilon_n} : n \in \mathbb{N}\}$ is a tight family of measures. By the Harnack inequality the family $\{\hat{\varrho}_z^{\varepsilon_n} : n \in \mathbb{N}\}$ is locally bounded, and locally Hölder equicontinuous, and the same of course applies to $\{\mathring{\varrho}_z^{\varepsilon_n} : n \in \mathbb{N}\}$. Moreover, the tightness of $\{\mathring{\eta}_z^{\varepsilon_n} : n \in \mathbb{N}\}$ implies the uniform integrability of $\{\mathring{\varrho}_z^{\varepsilon_n} : n \in \mathbb{N}\}$. Select any subsequence, also denoted by $\{\varepsilon_n\}$, along which $\mathring{\varrho}_z^{\varepsilon_n}$ converges locally uniformly, and denote the limit by $\bar{\varrho}_z$. By uniform integrability, $\mathring{\varrho}_z^{\varepsilon_n}$ also converges in $L^1(\mathbb{R}^d)$, as $n \to \infty$, and hence $\int_{\mathbb{R}^d} \bar{\varrho}_z(x) \, \mathrm{d}x = 1$. Therefore, $\bar{\eta}_z(\mathrm{d}x) := \bar{\varrho}_z(x) \, \mathrm{d}x$ is a probability measure. Let f be a smooth function with compact support, and $\bar{\mathcal{L}}_z$ be as in (5.7). Then

$$\begin{aligned} &\left| \int_{\mathbb{R}^d} \widehat{\mathcal{L}}_z^{\varepsilon_n} f(x) \mathring{\varrho}_z^{\varepsilon_n}(x) \, \mathrm{d}x - \int_{\mathbb{R}^d} \bar{\mathcal{L}}_z f(x) \bar{\varrho}_z(x) \, \mathrm{d}x \right| \\ &\quad \le \left| \int_{\mathbb{R}^d} \widehat{\mathcal{L}}_z^{\varepsilon_n} f(x) \big(\mathring{\varrho}_z^{\varepsilon_n}(x) - \bar{\varrho}_z(x) \big) \, \mathrm{d}x \right| \\ &\qquad + \left| \int_{\mathbb{R}^d} \big(\widehat{\mathcal{L}}_z^{\varepsilon_n} f(x) - \bar{\mathcal{L}}_z f(x) \big) \bar{\varrho}_z(x) \, \mathrm{d}x \right|. \end{aligned} \tag{5.10}$$

Since $\mathring{\varrho}_z^{\varepsilon_n} \to \bar{\varrho}_z$ in $L^1(\mathbb{R}^d)$, the first term on the right-hand side of (5.10) tends to 0 as $n \to \infty$. Similarly, since $\widehat{m}_z^{\varepsilon_n}(x) \to M_z x$ and $\nabla \widehat{V}_z^{\varepsilon_n} \to \nabla \bar{V}_z$ uniformly on compact subsets of $\mathbb{R}^d$, the second term also tends to 0. Since $\hat{\eta}_z^\varepsilon$ is an invariant probability measure of (5.9), by the definition of $\mathring{\varrho}_z^{\varepsilon_n}$ we have $\int_{\mathbb{R}^d} \widehat{\mathcal{L}}_z^{\varepsilon_n} f(x) \mathring{\varrho}_z^{\varepsilon_n}(x) \, \mathrm{d}x = 0$, for all large enough n, which implies that $\int_{\mathbb{R}^d} \bar{\mathcal{L}}_z f(x) \bar{\varrho}_z(x) \, \mathrm{d}x = 0$. Hence, $\bar{\eta}_z$ is an infinitesimal invariant probability measure of (5.6), and since the diffusion is regular, it is also an invariant probability measure. This proves part (a).

Since the diffusion in (5.6) has an invariant probability measure, it follows that it is positive recurrent. By Lemma 4.1, we have

$$\sup_{\varepsilon \in (0,1)} \int_{\{\varepsilon^\nu x + z \in \mathcal{B}_z\}} |x|^2 \hat{\eta}_z^\varepsilon(\mathrm{d}x) < \infty,$$

which implies by Fatou's lemma that $\int_{\mathbb{R}^d} |x|^2 \bar{\eta}_z(\mathrm{d}x) < \infty$. In addition, by Theorem 1.4 and Theorem 1.19(c) we must have $\bar{\beta}_* - \ell(z) = \Lambda^+(M_z)$. This completes the proof of parts (b) and (c).

By part (c) and Theorem 1.19(c), the solution of (5.1) is unique and is given by (5.8). That $\bar{\varrho}_z$ is Gaussian with covariance matrix $\widehat{\Sigma}_z$ follows by the second equation in (1.17). This proves part (d).

Since $\bar{V}_z$ has at most quadratic growth by (5.5), we have

$$\int_{\mathbb{R}^d} \bigl|\bar{V}_z(x)\bigr| \bar{\eta}_z(\mathrm{d}x) < \infty.$$

Therefore, with $\overline{\mathbb{E}}_x$ denoting the expectation operator for the process governed by (5.6), it is the case that $\overline{\mathbb{E}}_x[\bar{V}_z(X_t)]$ converges as $t \to \infty$ [Ichihara (2012), Theorem 4.12]. Integrating both sides of (5.1) with respect to $\bar{\eta}_z$, we deduce that

$$\int_{\mathbb{R}^d} \frac{1}{2} \bigl|\nabla \bar{V}_z(x)\bigr|^2 \bar{\eta}_z(\mathrm{d}x) = \bar{\beta}_* - \ell(z). \tag{5.11}$$

Using Fatou's lemma, we obtain by part (d) that

$$\begin{aligned}
&\liminf_{\varepsilon_n \searrow 0} \int_{\mathcal{B}_z} \mathcal{R}\bigl[v_*^{\varepsilon_n}\bigr](x) \eta_*^{\varepsilon_n}(\mathrm{d}x) \\
&\quad = \liminf_{\varepsilon_n \searrow 0} \int_{\{\varepsilon^\nu x + z \in \mathcal{B}_z\}} \Bigl(\widehat{\ell}_z^{\varepsilon_n}(x) + \frac{1}{2} \bigl|\nabla \widehat{V}_z^{\varepsilon_n}(x)\bigr|^2 \Bigr) \hat{\eta}_z^{\varepsilon_n}(\mathrm{d}x) \\
&\quad \geq \lim_{R \to \infty} \liminf_{\varepsilon_n \searrow 0} \int_{\{|x| \leq R\}} \Bigl(\widehat{\ell}_z^{\varepsilon_n}(x) + \frac{1}{2} \bigl|\nabla \widehat{V}_z^{\varepsilon_n}(x)\bigr|^2 \Bigr) \eta_*^{\varepsilon_n}(\mathcal{B}_z) \mathring{\eta}_z^{\varepsilon_n}(\mathrm{d}x) \\
&\quad \geq \theta_z \bigl(\Lambda^+(M_z) + \ell(z)\bigr),
\end{aligned}$$

where in the second inequality we use (5.11), along with the hypothesis that $\eta_*^{\varepsilon_n}(\mathcal{B}_z) \to \theta_z > 0$. This proves part (e) and thus completes the proof. □

Part of the statement in Theorem 1.11(iii) follows from the following result.

THEOREM 5.4. *Recall the definition of* $\mathfrak{J}_{\mathrm{c}}$ *from Theorem* 1.11. *We assume* $\nu = 1$. *Then, it holds that* $\lim_{\varepsilon \searrow 0} \beta_*^\varepsilon = \mathfrak{J}_{\mathrm{c}}$. *In addition,* $\bar{\beta}_*$ *in* (5.1) *equals* $\mathfrak{J}_{\mathrm{c}}$. *Moreover, for any* $r > 0$ *we have*

$$\lim_{\varepsilon \searrow 0} \eta_*^\varepsilon\bigl(B_r^c(\mathcal{Z}_{\mathrm{c}})\bigr) = 0 \quad \textit{and} \quad \lim_{\varepsilon \searrow 0} \int_{B_r^c(\mathcal{Z}_{\mathrm{c}})} \bigl|v_*^\varepsilon(x)\bigr|^2 \eta_*^\varepsilon(\mathrm{d}x) = 0. \tag{5.12}$$

PROOF. Since the collection $\{\mathcal{B}_z\}$ used in Theorem 5.3 was arbitrary, without loss of generality, we may let $\mathcal{B}_z = B_r(z)$. Let $\varepsilon_n \searrow 0$ be any sequence such that $\eta_*^{\varepsilon_n}(B_r(z)) \to \theta_z$ as $n \to \infty$, for all $z \in \mathcal{S}$, and define $\mathcal{S}_{\mathrm{o}} := \{z \in \mathcal{S} \colon \theta_z > 0\}$. Since

$\mathcal{S}$ is stochastically stable as shown in Theorem 1.11, we have $\sum_{z\in\mathcal{S}_{\mathrm{o}}}\theta_z = 1$. By Theorem 5.3(e) we have

$$\begin{aligned}\liminf_{n\to\infty}\beta^{\varepsilon_n}_* &\geq \sum_{z\in\mathcal{S}_{\mathrm{o}}}\int_{B_r(z)}\Big(\ell(x)+\frac{1}{2}\big|v^{\varepsilon_n}_*(x)\big|^2\Big)\eta^{\varepsilon_n}_*(\mathrm{d}x)\\ &\geq \sum_{z\in\mathcal{S}_{\mathrm{o}}}\theta_z\big(\ell(z)+\Lambda^+\big(Dm(z)\big)\big)\geq \mathfrak{J}_{\mathrm{c}}.\end{aligned}\tag{5.13}$$

Since $\limsup_{\varepsilon\searrow 0}\beta^{\varepsilon}_*\leq\mathfrak{J}_{\mathrm{c}}$ by Lemma 3.3, (5.13) implies that $\lim_{\varepsilon\searrow 0}\beta^{\varepsilon}_*=\mathfrak{J}_{\mathrm{c}}$. By Lemma 5.2, we have $\liminf_{\varepsilon\searrow 0}\beta^{\varepsilon}_*\leq\bar\beta_*$, and $\bar\beta_*\leq\mathfrak{J}_{\mathrm{c}}$ by (5.3). Therefore, $\bar\beta_*=\mathfrak{J}_{\mathrm{c}}$.

Given any sequence $\varepsilon_n\searrow 0$, we can extract a subsequence also denoted by $\{\varepsilon_n\}$ along which $\lim_{n\to\infty}\eta^{\varepsilon_n}_*(B_r(z))\to\theta_z$ for all $z\in\mathcal{S}$. Then (5.13) holds. In addition, by Proposition 4.5, we have $\int_{B_r^c(z)}\ell(x)\eta^{\varepsilon}_*(\mathrm{d}x)\to 0$ as $\varepsilon\searrow 0$. It is then clear that both assertions in (5.12) follow by (5.13). □

It is interesting to note that, even if $\lim_{\varepsilon\to 0}\eta^{\varepsilon_n}_*(\mathcal{B}_z)=0$, equation (5.8) still holds for any $z\in\mathcal{Z}_{\mathrm{c}}$. This is part of the corollary that follows.

COROLLARY 5.5. *Suppose $\nu=1$. Then for any $z\in\mathcal{Z}_{\mathrm{c}}$, we have*

$$\widehat{V}^{\varepsilon}_z(x)-\widehat{V}^{\varepsilon}_z(z)\xrightarrow[\varepsilon\searrow 0]{}\frac{1}{2}\langle x,\widehat{Q}_z x\rangle,$$

uniformly on compact sets. In addition, unless $z\in\mathcal{Z}_{\mathrm{c}}$, the family $\{\mathring{\eta}^{\varepsilon}_z:\varepsilon\in(0,1)\}$ is not tight.

PROOF. Since $\bar\beta_*$ in (5.1) equals $\mathfrak{J}_{\mathrm{c}}$ by Theorem 5.4, then, provided $z\in\mathcal{Z}_{\mathrm{c}}$, the right-hand side of (5.1) equals $\Lambda^+(M_z)$. The first assertion then follows by Theorem 1.19(c).

If the family $\{\mathring{\eta}^{\varepsilon}_z:\varepsilon\in(0,1)\}$ is tight, then it follows from the proof of Theorem 5.3 that the diffusion limit in (5.6) is positive recurrent. However, if $z\notin\mathcal{Z}_{\mathrm{c}}$, then $\bar\beta_*-\ell(z)=\mathfrak{J}_c-\ell(z)<\Lambda^+(M_z)$, and by the results of Theorem 1.4 and Theorem 1.19(c), the diffusion in (5.6) has to be transient. Therefore, $\{\mathring{\eta}^{\varepsilon}_z\}$ cannot be tight. □

REMARK 5.6. It is worth examining the diffusion in (5.6) in the context of Example 1.14. Consider the example with the first set of data, and let $c=5$. Then $\mathfrak{S}=\{0\}$ and $\mathfrak{J}_{\mathrm{c}}=2$. Thus, for $z=0$, we have $\bar V_z=\bar V_0=2x^2$, and the drift in (5.6) equals $-2\bar X_t$. For $z=-1$, we have $\ell(-1)=5$, $Dm(-1)=-3$, and direct substitution shows that $\bar V_{-1}=-3x^2$ solves (5.1). The associated diffusion in (5.6) has drift $3\bar X_t$, and thus it is transient.

5.2. *Convergence to a Gaussian in the subcritical/supercritical regime.* We return to the analysis of the subcritical and supercritical regimes in order to determine the asymptotic behavior of the density of the optimal stationary distribution in the vicinity of the stochastically stable set. In these regimes there are two scales. If we center the coordinates around a point in $\mathfrak{S}$, then we have $V^\varepsilon(x) \in \mathcal{O}(\varepsilon^{-2}|x|^2)$, and $-\log \varrho_*^\varepsilon(x) \in \mathcal{O}(\varepsilon^{-2\nu}|x|^2)$. To avoid this incompatibility we use the function $\check{V}_z(x) = \varepsilon^{2(1-\nu)} V^\varepsilon(\varepsilon^\nu x)$ in the analysis, which scales correctly in space for all ν. We have the following result.

THEOREM 5.7. *Assume $\nu \in (0,2)$ and let $\{\mathcal{B}_z : z \in \mathcal{S}\}$ be as in Definition 5.1. The following hold.*

(a) *Suppose that for some $z \in \mathcal{S}$ and a sequence $\varepsilon_n \searrow 0$ it holds that $\liminf_{n\to\infty} \eta_*^{\varepsilon_n}(\mathcal{B}_z) > 0$. Then the density $\mathring{\varrho}_z^{\varepsilon_n}$ in Definition 5.1 converges as $n \to \infty$ (uniformly on compact sets) to the density of a Gaussian with mean 0 and covariance matrix $\widehat{\Sigma}_z$ given in (1.17).*

(b) *If $\nu \in (1,2)$ and $z \in \mathcal{S} \setminus \widetilde{\mathcal{Z}}$, then $\lim_{\varepsilon\searrow 0} \eta_*^\varepsilon(\mathcal{B}_z) = 0$.*

PROOF. The proof closely follows those of Lemma 5.2 and Theorem 5.3. Only the scaling differs. We summarize the essential steps.

First, suppose $\nu < 1$. Since $\liminf_{n\to\infty} \eta_*^{\varepsilon_n}(\mathcal{B}_z) > 0$, then necessarily $z \in \mathcal{S}_s$ by Lemma 3.6. We scale the space as $\frac{1}{\varepsilon^\nu}$, and use (4.15) which we write again here as

$$\begin{aligned} &\frac{1}{2}\Delta \check{V}_z^\varepsilon(x) + \langle \widehat{m}_z^\varepsilon(x), \nabla \check{V}_z^\varepsilon(x)\rangle - \frac{1}{2}|\nabla \check{V}_z^\varepsilon(x)|^2 \\ &\qquad = \varepsilon^{2(1-\nu)}\big(\beta_*^\varepsilon - \widehat{\ell}_z^\varepsilon(x)\big). \end{aligned} \tag{5.14}$$

By Lemma 4.4, $\nabla \check{V}_z^\varepsilon = \varepsilon^{2(1-\nu)} \nabla \widehat{V}_z^\varepsilon$ is locally bounded and has at most linear growth. We write (5.14) as a HJB equation in the form

$$\begin{aligned} &\frac{1}{2}\Delta \check{V}_z^\varepsilon(x) + \min_{\check{u}\in\mathbb{R}^d}\Big[\langle \widehat{m}_z^\varepsilon(x) + \check{u}, \nabla \check{V}_z^\varepsilon(x)\rangle + \frac{1}{2}|\check{u}|^2\Big] \\ &\qquad = \varepsilon^{2(1-\nu)}\big(\beta_*^\varepsilon - \widehat{\ell}_z^\varepsilon(x)\big). \end{aligned} \tag{5.15}$$

The associated scaled controlled diffusion is

$$\mathrm{d}\widehat{X}_t = \big(\widehat{m}_z^\varepsilon(\widehat{X}_t) - \check{U}_t\big)\,\mathrm{d}t + \mathrm{d}\widehat{W}_t. \tag{5.16}$$

Taking limits in (5.15) along some subsequence $\varepsilon_n \searrow 0$, we obtain a function $\bar{V}_z \in \mathcal{C}^2(\mathbb{R}^d)$ of at most quadratic growth, satisfying

$$\frac{1}{2}\Delta \bar{V}_z(x) + \min_{\bar{u}\in\mathbb{R}^d}\Big[\langle M_z x + \bar{u}, \nabla \bar{V}_z(x)\rangle + \frac{1}{2}|\bar{u}|^2\Big] = 0. \tag{5.17}$$

The associated diffusion limit is

$$\mathrm{d}\bar{X}_t = \big(M_z \bar{X}_t - \nabla \bar{V}_z(\bar{X}_t)\big)\,\mathrm{d}t + \mathrm{d}\bar{W}_t. \tag{5.18}$$

As in Section 5.1, $\hat{\eta}^\varepsilon_*$ denotes the invariant probability measure of (5.16) under the control $\check{U}_t = -\nabla \check{V}^\varepsilon_z(X_t)$, and $\hat{\varrho}^\varepsilon_*$ its density. Following the proof of Theorem 5.3, and using Lemma 4.1, we deduce that the density $\mathring{\varrho}^\varepsilon_z$ in Definition 5.1 converges as $\varepsilon_n \searrow 0$ to the density $\bar{\varrho}_z$ of the invariant probability measure of (5.18). However, since M_z is Hurwitz, then $\Lambda^+(M_z) = 0$, and by Theorem 1.19 we obtain $\bar{V}_z \equiv 0$. So in this case (5.17) is trivial, and the covariance matrix $\widehat{\Sigma}_z$ of the Gaussian is the solution of (1.17) with $\widehat{Q}_z = 0$.

Next, we assume $\nu \in (1,2)$, and we use the same scaling and definitions as for the subcritical regime, except that $z \in \mathcal{Z}$. It is clear that

$$\varepsilon^{2(1-\nu)}\bigl(\widehat{\ell}^\varepsilon_z(x) - \ell(z)\bigr) \le C_\ell \varepsilon^{2(1-\nu)}\varepsilon^\nu |x| \xrightarrow[\varepsilon \searrow 0]{} 0,$$

where C_ℓ denotes a Lipschitz constant of ℓ. We have $\mathcal{G}^\varepsilon_* \in \mathcal{O}(\varepsilon^{2\nu-2})$ by Corollary 4.2. In addition, by Lemma 4.1 and Proposition 4.5, and since $\nu \in (1,2)$, we obtain

$$\left| \int_{\mathbb{R}^d} \ell \, \mathrm{d}\eta^\varepsilon_* - \ell(z) \right| \in \mathcal{O}\bigl(\varepsilon^\nu\bigr).$$

It follows that the constants $\varepsilon^{2(1-\nu)}(\beta^\varepsilon_* - \ell(z))$ are bounded, uniformly in $\varepsilon \in (0,1)$. Therefore, as argued in the proof of Theorem 5.3, for every sequence $\varepsilon_n \searrow 0$, there exists a subsequence, also denoted as $\{\varepsilon_n\}$ along which $\varepsilon_n^{2(1-\nu)}(\beta^\varepsilon_* - \ell(x))$ converges to a constant $\hat{\beta}$, and $\check{V}^\varepsilon_z(\cdot) - \check{V}^\varepsilon_z(z)$ converges to some $\bar{V}_z \in \mathcal{C}^2(\mathbb{R}^d)$, uniformly on compact sets. Taking limits in (5.15) along this subsequence, we obtain

$$\frac{1}{2}\Delta \bar{V}_z(x) + \min_{\bar{u} \in \mathbb{R}^d}\left[\bigl\langle M_z x + \bar{u}, \nabla \bar{V}_z(x)\bigr\rangle + \frac{1}{2}|\bar{u}|^2\right] = \hat{\beta}. \tag{5.19}$$

Recall the notation $\widetilde{\mathcal{Z}}$ and $\widetilde{\mathfrak{J}}$ in Definition 1.10. By Lemma 3.3, we have

$$\hat{\beta} \le \widetilde{\mathfrak{J}} = \min_{z \in \mathcal{Z}} \Lambda^+\bigl(Dm(z)\bigr). \tag{5.20}$$

Following exactly the same steps as in the proof of Theorem 5.3, we deduce that the diffusion in (5.18) is positive recurrent, with an invariant probability measure $\bar{\eta}_z$ that has finite second moments, and that the density $\mathring{\varrho}^\varepsilon_z$ in Definition 5.1 converges as $\varepsilon_n \searrow 0$ to the density $\bar{\varrho}_z$ of $\bar{\eta}_z$. Therefore,

$$\Lambda^+\bigl(Dm(z)\bigr) = \hat{\beta} \tag{5.21}$$

by Theorem 1.19(c). Thus we obtain $\hat{\beta} = \widetilde{\mathfrak{J}} = \Lambda^+(Dm(z))$ by (5.20)–(5.21). This shows that unless $z \in \widetilde{\mathcal{Z}}$, the hypothesis $\liminf_{n\to\infty} \eta^{\varepsilon_n}_*(\mathcal{B}_z) > 0$ cannot hold, thus establishing part (b) of the theorem.

With $z \in \widetilde{\mathcal{Z}}$, and $\hat{\beta} = \widetilde{\mathfrak{J}}$, equation (5.19) has a unique solution by Theorem 1.19(c), and we obtain $\bar{V}_z(x) = \frac{1}{2}\langle x, \widehat{Q}_z x\rangle$, and that $\bar{\varrho}_z$ is the density of a Gaussian with mean 0 and covariance matrix $\widehat{\Sigma}_z$, where $(\widehat{Q}_z, \widehat{\Sigma}_z)$ is the pair of matrices which solve (1.17). This completes the proof. $\square$

6. Concluding remarks. In general, Morse–Smale flows may contain hyperbolic closed orbits, and it would be desirable to extend the results of the paper accordingly. An energy function $\mathcal{V}$ as in Theorem 2.2 may be constructed to account for critical elements that are closed orbits Meyer (1968), Smale (1961). Note that under the control used in Remark 3.4, the optimal stationary distribution concentrates on the minimum of $\mathcal{V}$. In the case that $z \in \mathbb{R}^d$ belongs to a stable periodic orbit with period T_0, we can construct $\mathcal{V}$ so that it attains its minimum on this closed orbit. In this manner, if ϕ_t denotes the flow of the vector field m, it then follows by (3.8) that under the control used in Remark 3.4, we obtain

$$\int_{\mathbb{R}^d} \ell(x)\mu^{\varepsilon}(\mathrm{d}x) \xrightarrow[\varepsilon \searrow 0]{} \frac{1}{T_0}\int_0^{T_0} \ell\big(\phi_t(z)\big)\,\mathrm{d}t.$$

The same can be accomplished for the subcritical regime, by modifying the proof of Lemma 3.5, and using instead the approach in Remark 3.4. We leave it up to the reader to verify that Lemma 3.1 still holds if the set of critical elements $\mathcal{S}$ contains hyperbolic closed orbits. Let us define

$$\mathring{\ell}(z) := \frac{1}{T_0}\int_0^{T_0} \ell\big(\phi_t(z)\big)\,\mathrm{d}t,$$

when z belongs to a closed orbit, and $\mathring{\ell}(z) = \ell(z)$, when $m(z) = 0$. Then, provided $\operatorname{Arg\,min}_{z\in\mathcal{S}} \mathring{\ell}(z)$ contains only stable critical elements, then the support of the limit of the optimal stationary distribution lies in $\mathcal{S}_{\mathrm{s}}$, and this is true in any of the three regimes. However, the full analysis when unstable closed orbits are involved seems to be more difficult.

Acknowledgments. The authors are indebted to the anonymous referees for their constructive comments and suggestions. This work was initiated during Vivek Borkar's visit to the Department of Electrical Engineering, Technion, supported by Technion. Thanks are due to Professor Rami Atar for suggesting the problem as well as for valuable discussions.

SUPPLEMENTARY MATERIAL

Proofs of Lemma 1.3, Theorem 1.4, Lemma 1.17, and Theorem 1.19 (DOI: 10.1214/17-AOP1238SUPP; .pdf). This supplement contains the proofs of Lemma 1.3 and Theorem 1.4 in Section 1.3, and Lemma 1.17 and Theorem 1.19 in Section 1.5.

REFERENCES

ARAPOSTATHIS, A., BISWAS, A. and BORKAR, V. S. (2018). Supplement to "Controlled equilibrium selection in stochastically perturbed dynamics." DOI:10.1214/17-AOP1238SUPP.

ARAPOSTATHIS, A., BORKAR, V. S. and GHOSH, M. K. (2012). *Ergodic Control of Diffusion Processes*. *Encyclopedia of Mathematics and Its Applications* **143**. Cambridge Univ. Press, Cambridge. MR2884272

BENSOUSSAN, A. and FREHSE, J. (1992). On Bellman equations of ergodic control in $\mathbf{R}^n$. *J. Reine Angew. Math.* **429** 125–160. MR1173120

BENSOUSSAN, A. and FREHSE, J. (2002). *Regularity Results for Nonlinear Elliptic Systems and Applications*. *Applied Mathematical Sciences* **151**. Springer, Berlin.

BENZI, R., PARISI, G., SUTERA, A. and VULPIANI, A. (1983). A theory of stochastic resonance in climatic change. *SIAM J. Appl. Math.* **43** 565–478.

BERGLUND, N. and GENTZ, B. (2002). Metastability in simple climate models: Pathwise analysis of slowly driven Langevin equations. *Stoch. Dyn.* **2** 327–356.

BERGLUND, N. and GENTZ, B. (2006). *Noise-Induced Phenomena in Slow-Fast Dynamical Systems. A Sample-Paths Approach*. Springer, London.

BHATT, A. G. and BORKAR, V. S. (1996). Occupation measures for controlled Markov processes: Characterization and optimality. *Ann. Probab.* **24** 1531–1562. MR1411505

BISWAS, A. and BORKAR, V. S. (2009). Small noise asymptotics for invariant densities for a class of diffusions: A control theoretic view. *J. Math. Anal. Appl.* **360** 476–484. Erratum at arXiv:1107.2277.

BOBROVSKY, B. Z., ZAKAI, M. M. and ZEITOUNI, O. (1988). Error bounds for the nonlinear filtering of signals with small diffusion coefficients. *IEEE Trans. Inform. Theory* **34** 710–721.

BOGACHEV, V. I., RÖCKNER, M. and SHAPOSHNIKOV, S. V. (2012). On positive and probability solutions of the stationary Fokker–Planck–Kolmogorov equation. *Dokl. Akad. Nauk* **444** 245–249.

BORKAR, V. S. (1989). *Optimal Control of Diffusion Processes*. *Pitman Research Notes in Mathematics Series* **203**. Longman Scientific & Technical, Harlow. MR1005532

CERRAI, S. and RÖCKNER, M. (2005). Large deviations for invariant measures of stochastic reaction-diffusion systems with multiplicative noise and non-Lipschitz reaction term. *Ann. Inst. Henri Poincaré Probab. Stat.* **41** 69–105. MR2110447

DAY, M. V. (1987). Recent progress on the small parameter exit problem. *Stochastics* **20** 121–150.

FENG, J., FORDE, M. and FOUQUE, J.-P. (2010). Short-maturity asymptotics for a fast mean-reverting Heston stochastic volatility model. *SIAM J. Financial Math.* **1** 126–141. MR2592567

FREIDLIN, M. I. and WENTZELL, A. D. (1998). *Random Perturbations of Dynamical Systems*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* **260**. Springer, New York.

HERRMANN, S., IMKELLER, P., PAVLYUKEVICH, I. and PEITHMANN, D. (2014). *Stochastic Resonance*. *Mathematical Surveys and Monographs* **194**. Amer. Math. Soc., Providence, RI.

ICHIHARA, N. (2011). Recurrence and transience of optimal feedback processes associated with Bellman equations of ergodic type. *SIAM J. Control Optim.* **49** 1938–1960.

ICHIHARA, N. (2012). Large time asymptotic problems for optimal stochastic control with superlinear cost. *Stochastic Process. Appl.* **122** 1248–1275.

ICHIHARA, N. (2015). The generalized principal eigenvalue for Hamilton–Jacobi–Bellman equations of ergodic type. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **32** 623–650.

ICHIHARA, N. and SHEU, S. J. (2013). Large time behavior of solutions of Hamilton–Jacobi–Bellman equations with quadratic nonlinearity in gradients. *SIAM J. Math. Anal.* **45** 279–306.

KRYLOV, N. V. and RÖCKNER, M. (2005). Strong solutions of stochastic equations with singular time dependent drift. *Probab. Theory Related Fields* **131** 154–196.

KUČERA, V. (1972). A contribution to matrix quadratic equations. *IEEE Trans. Automat. Control* **AC-17** 344–347.

KURTZ, T. G. and STOCKBRIDGE, R. H. (1998). Existence of Markov controls and characterization of optimal Markov controls. *SIAM J. Control Optim.* **36** 609–653.

LINDNER, J. F., BENNETT, M. and WIESENFELD, K. (2006). Potential energy landscape and finite-state models of array-enhanced stochastic resonance. *Phys. Rev. E* **73** 031107.

MÅRTENSSON, K. (1971). On the matrix Riccati equation. *Inform. Sci.* **3** 17–49.

METAFUNE, G., PALLARA, D. and RHANDI, A. (2005). Global properties of invariant measures. *J. Funct. Anal.* **223** 396–424.

MEYER, K. R. (1968). Energy functions for Morse Smale systems. *Amer. J. Math*. **90** 1031–1040. MR0239220

MOSS, F. (1994). Stochastic resonance: From the ice ages to the monkey's ear. In *Contemporary Problems in Statistical Physics* (G. H. Weiss, ed.) 205–253 5. SIAM, Philadelphia.

OLIVIERI, E. and VARES, M. E. (2005). *Large Deviations and Metastability*. *Encyclopedia of Mathematics and Its Applications* **100**. Cambridge Univ. Press, Cambridge.

REPPERGER, D. W. and FARRIS, K. A. (2010). Stochastic resonance—A nonlinear control theory interpretation. *Internat. J. Systems Sci*. **41** 897–907.

RUSSELL, D. F., WILKENS, L. A. and MOSS, F. (1999). Use of behavioural stochastic resonance by paddle fish for feeding. *Nature* **402** 291–294.

SCHUSS, Z. (1980). *Theory and Applications of Stochastic Differential Equations*. Wiley, New York.

SHEU, S. J. (1986). Asymptotic behavior of the invariant density of a diffusion Markov process with small diffusion. *SIAM J. Math. Anal*. **17** 451–460.

SMALE, S. (1961). On gradient dynamical systems. *Ann. of Math*. (2) **74** 199–206. MR0133139

WILLEMS, J. C. (1971). Least squares stationary optimal control and the algebraic Riccati equation. *IEEE Trans. Automat. Control* **AC-16** 621–634.

WU, X., JIANG, Z. P., REPPERGER, D. W. and GUO, Y. (2006). Enhancement of stochastic resonance using optimization theory. *Commun. Inf. Syst*. **6** 1–18.

YANG, Y., JIANG, Z., XU, B. and REPPERGER, D. W. (2009). An investigation of two-dimensional parameter-induced stochastic resonance and applications in nonlinear image processing. *J. Phys. A* **42** 145207, 9. MR2539249

ZEITOUNI, O. and ZAKAI, M. (1992). On the optimal tracking problem. *SIAM J. Control Optim*. **30** 426–439. MR1149077

A. ARAPOSTATHIS
DEPARTMENT OF ELECTRICAL
AND COMPUTER ENGINEERING
UNIVERSITY OF TEXAS AT AUSTIN
2501 SPEEDWAY, EER 7.824
AUSTIN, TEXAS 78712
USA
E-MAIL: ari@ece.utexas.edu

A. BISWAS
INDIAN INSTITUTE OF SCIENCE EDUCATION
AND RESEARCH
DR. HOMI BHABHA ROAD
PUNE 411008
INDIA
E-MAIL: anup@iiserpune.ac.in

V. S. BORKAR
DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY
POWAI, MUMBAI
INDIA
E-MAIL: borkar@ee.iitb.ac.in