The Annals of Probability
1996, Vol. 24, No. 3, 1239–1268

HARMONIC COORDINATES FOR DIFFUSIONS IN THE PLANE

BY G. KERSTING

University of Frankfurt

For a class of diffusions X in the plane we construct a global system of coordinates $u(x)$, such that $u(x)$ is close to x at infinity and $u(X)$ is a local martingale. Such coordinates are useful for the study of the long term behaviour of X. The construction uses probabilistic methods, in particular a coupling for general diffusions in the plane.

1. Introduction and main result. In this paper we consider coordinate systems for diffusions in the plane, which in dimension 1 usually are called natural scales. We do not use this term here. It is doubtful that in a higher-dimensional setting such coordinates will always exist; also they serve quite different purposes. In dimension 1 these coordinate systems are particularly useful to construct regular diffusion processes in full generality (see [21], Chapter 5.7 for an account), which in higher dimensions is out of reach. We have quite a different kind of question in mind: Imagine that random noise is added to some deterministic dynamical system: what is the effect? Is it negligible or will there be a drastic change in its long term behaviour? For such a kind of problem harmonic coordinate systems turn out to be rather useful. We shall discuss this in detail in Section 2. For the moment, let us just explain what we mean by harmonic coordinates. Some notational conventions: Vectors are always row vectors, with Euclidean norm $|\ |$. $\|\ \|$ is the usual matrix norm. ξ^* is the transpose of ξ. The center dot $(\cdot)$ denotes matrix multiplication.

Consider a diffusion $X = (X_t)$ in $\mathbb{R}^d$, $d \geq 1$, given by the Itô equation

$$dX_t = b(X_t)\,dt + \sigma(X_t) \cdot dW_t, \qquad t \geq 0,$$

where $b(x) = (b_1(x), \dots, b_d(x))^*$ is in $\mathbb{R}^d$, $\sigma(x)$ denotes a $d \times n$ matrix and $W = (W_t)$ is an n-dimensional standard Brownian motion. Let

$$a(x) = \big(a_{ij}(x)\big) = \sigma(x) \cdot \sigma(x)^*$$

Received November 1993; revised August 1995.

AMS 1991 *subject classifications*. Primary 60J60; secondary 60H30, 57R50, 58G32.

Key words and phrases. Diffusion processes, dynamical systems, harmonic mappings, harmonic coordinates, coupling methods, Poisson equation in unbounded domains, a priori estimates.

be the corresponding diffusion matrix at $x \in \mathbb{R}^d$ and let

$$L = \frac{1}{2}\sum_{i,j} a_{ij}(x)\frac{\partial^2}{\partial x_i\,\partial x_j} + \sum_i b_i(x)\frac{\partial}{\partial x_i}$$

be the infinitesimal generator. Then a bijective C^2-mapping $u\colon \mathbb{R}^d \to \mathbb{R}^d$ is called *harmonic system of coordinates* if

$$Lu \equiv 0.$$

(L is applied componentwise.) The main effect of a harmonic change of the state space's coordinates is that the drift component is completely removed from our diffusion process. This is a consequence of Itô's formula, which in the case of $Lu \equiv 0$ reads

$$du(X_t) = Du(X_t)\cdot\sigma(X_t)\cdot dW_t.$$

Here $Du(x)$ denotes the Jacobian matrix of u in x.

When do harmonic coordinates exist? For our result some smoothness of $b(x)$ and $a(x)$ is required. The following assumption is convenient:

ASSUMPTION A. $a_{ij}(x)$ and $b_i(x)$ are locally Hölder continuous for all i, j.

[A function f is called locally Hölder continuous if for any compact set K there are $\alpha, \beta > 0$ such that $|f(x) - f(y)| \le \beta|x - y|^\alpha$ for all $x, y \in K$.] Our main assumption concerns the diffusion matrix. We require uniform ellipticity:

ASSUMPTION E. There is a number $C > 0$ such that for any $x \in \mathbb{R}^d$ and any vectors ξ and η of length 1,

$$0 < \xi^*\cdot a(x)\cdot\xi \le C\eta^*\cdot a(x)\cdot\eta.$$

In other words, $0 < \lambda_{\max}(x) \le C\lambda_{\min}(x)$ for the maximal and minimal eigenvalue of $a(x)$.

In this paper we focus our attention on dimension $d = 2$ and on systems of harmonic coordinates possessing some additional favorable properties, which will play a crucial role in the next section.

THEOREM 1.1. *Let $d = 2$ and let Assumptions A and E be satisfied. If for some $\varepsilon > 0$,*

$$|x|^{1+\varepsilon}|b(x)| = O(\|a(x)\|) \tag{1.1}$$

as $|x| \to \infty$, then there exists a mapping $u\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $Lu \equiv 0$ and

$$|u(x) - x| = o(|x|)$$

as $|x| \to \infty$. u has the following properties:

(i) *u is a global C^2-diffeomorphism, a regular bijection from $\mathbb{R}^2$ onto $\mathbb{R}^2$.*
(ii) $Du(x) \to \mathrm{Id}$, *the identity matrix*, *as $|x| \to \infty$,*
(iii) *u is uniquely determined*, *up to an additional translation in $\mathbb{R}^2$.*

The following comments may help to judge this result:

1. For dimension $d = 1$ the corresponding statement is not entirely correct. Then the equation $Lu \equiv 0$ can be solved explicitly:
$$u'(x) = \exp\left(-\int^x \frac{2b(y)}{a(y)}\,dy\right).$$
Thus under (1.1) there are positive numbers α and β such that $u(x) \sim \alpha x$ for $x \to \infty$ and $u(x) \sim \beta x$ For $x \to -\infty$. $u(x) \sim x$ for $x \to \pm\infty$, however, can be achieved only in the exceptional case $\int_{-\infty}^{\infty}(b(y))/(a(y))\,dy = 0$. In dimension $d \geq 2$ this is different.
2. How good is condition (1.1)? It turns out that the conclusion of Theorem 1.1 may fail to be true if merely $|x| \cdot |b(x)| = O(\|a(x)\|)$ is assumed. This will be demonstrated in the next section: We shall construct a transient diffusion X_t with the property that $u(X_t)$ is recurrent for any mapping u as in Theorem 1.1. Of course this is not compatible with the requirement $|u(x) - x| = o(|x|)$, and consequently such a mapping cannot exist. The argument at the end of the present section also shows that $|x| \cdot |b(x)| = O(\|a(x)\|)$ is not good enough.
3. Applying the mapping u to the state space, the diffusion matrix will change from $a(x)$ to $Du(x) \cdot a(x) \cdot Du(x)^*$. Note, however, that we do not lose control over the diffusion matrix, since $Du(x) \to \mathrm{Id}$ at infinity. [It would be too much to try keep $a(x)$ fixed. For the complications arising, compare [2] in the case of Brownian motion.]

The main part of the paper (Sections 3–7 and Appendix) deals with the proof. Let us make some indications. To prove that u is a diffeomorphism, we use methods from topology; see Section 6. For the rest, in particular the construction of u, probabilistic methods turn out to be well suited. It is convenient to consider $v(x) = u(x) - x$. Then $Lu \equiv 0$ translates into the Poisson-type equation
$$Lv = -b$$
on $\mathbb{R}^d$, and $|u(x) - x| = o(|x|)$ changes into the growth restriction
$$|v(x)| = o(|x|) \tag{1.2}$$
at infinity. In the case of the Laplacian, $L = \Delta$, one could think of making use of the potential theoretic representation
$$\bar{v}(x) = \int_{\mathbb{R}^d} \Gamma(x, y) b(y)\,dy,$$
where $\Gamma(x, y)$ is the ordinary Green function. As is well known, $\bar{v}$ satisfies $L\bar{v} = -b$, provided the integral converges. However, then $|\bar{v}(x)| = o(1)$ for $|x| \to \infty$ such that this approach is not adequate for the growth restriction (1.2), since it will not cover those $v(x)$ which satisfy (1.2), but fail to fulfill $|v(x)| = o(1)$. Therefore, consider
$$v(x) = \int_{\mathbb{R}^d} (\Gamma(x, y) - \Gamma(0, y)) b(y)\,dy \tag{1.3}$$

instead, which now vanishes at $x = 0$ and no longer at ∞. This idea is developed in [16]; compare also [13] and [22].

For non-Laplacian L this method appears to be rather involved, because we can no longer control the Green function at infinity. Instead we shall use the well-known probabilistic representation

$$v(x) = E_x \int_0^\infty b(X_t)\,dt - E_0 \int_0^\infty b(X_t)\,dt \tag{1.4}$$

for solutions of $Lv = -b$. Again $v(x)$ is normalized such that $v(0) = 0$. The point is that we are able to give an exact definition for $v(x)$ even in situations where the expectations in (1.4) are no longer convergent. Roughly speaking we shall remove equal portions from both expectations. Details will be explained in Section 3.

It is certainly no surprise that a probabilistic coupling technique will play a central role in our construction. However, for our purposes common couplings for multidimensional diffusions, as those of Rogers and Lindvall or Chen and Li (cf. [20] and [4]) are by no means suited, even if we accept to replacement of Assumption A by an arbitrarily strong smoothness assumption. These couplings do not cover our general setting (compare also [11]). Our coupling makes use of an additional time change argument: it is based on the Stroock–Varadhan existence and uniqueness results for multidimensional SDE (see [23]). It has the favorable aspect that besides uniform ellipticity, it only requires continuity for the infinitesimal generator's coefficients. Its drawback is that it is confined to dimension 2; see Section 4. We will use the coupling also to derive estimates on the gradient (Section 5) and to prove uniqueness (Section 7).

Harmonic mappings have of course been considered before, and the question of existence (mainly between compact Riemannian manifolds) has drawn some attention. The extensive surveys in [7] contain further information. In probability harmonic mappings have been used rather efficiently for the study of two-dimensional Brownian motion (see Chapter 5 of [6] for an account). This has been pushed further by Kendall and others (see [14] and [15]), who focus on the interplay between geometry and Brownian motion on Riemannian manifolds, using harmonic mappings. On the other hand, it seems that not much is known about harmonic coordinates for general diffusions in dimension greater than 1.

Also the idea to remove the drift from a diffusion is certainly not new. One well-known method uses changes of measure and the Girsanov–Cameron–Martin formula. It is well suited to obtain weak solutions of SDE; see [23] for instance. (Compare also Kunita [17], who considers (nonharmonic) transformations of the state space in this context.) Zvonkin, on the other hand, eliminated the drift by transforming the state space (cf. [26]). His goal was the construction of strong solutions. (This was continued by Veretennikov [25].) In contrast to our approach, Zvonkin's transformation depends on the time parameter and is not harmonic in our sense. Both methods work only for finite periods of time, and they are not helpful for those kind of questions which we shall turn to in the next section.

Let us conclude the Introduction with some remarks on possible generalizations of our theorem. As we have argued, (1.1) with $\varepsilon = 0$ is not a sufficient assumption. Also $|x| \cdot |b(x)| = o(\|a(x)\|)$ is not enough. To see this suppose that

$$a(x) = \mathrm{Id}$$

and

$$b(x) = \beta(|x|)\frac{x}{|x|}$$

for some real function $\beta(r)$, $r > 0$. Under such symmetries harmonic coordinates have to be of the form

$$u(x) = f(|x|) \cdot x + c.$$

[This is due to uniqueness of $u(x)$; see Section 7.] It is not difficult to show that $Lu \equiv 0$ is then equivalent to

$$rf''(r) + (d + 1 + 2r\beta(r))f'(r) + 2\beta(r)f(r) = 0$$

for $r > 0$. We are interested in the situation where $\beta(r) = o(r^{-1})$ and $f(r) \to 1$ as $r \to \infty$. This implies that $f'(r)$ and $rf''(r)$ are finitely integrable at infinity. From the above equation we see that, in general, we can expect a solution $u(x)$ with the desired properties only if $\beta(r)$ is finitely integrable too. Now this seems to be the right assumption here. There is a lot of evidence for the following more general conjecture:

Conjecture 1.2. Let $\mathrm{d} \geq 2$ and let Assumption E be satisfied. Furthermore let $\varphi(\mathrm{r})$, $\mathrm{r} \geq 0$, be a decreasing function such that $\int_0^\infty \varphi(\mathrm{r})\,\mathrm{dr} < \infty$. If at infinity

$$|b(x)| = O(\varphi(|x|) \cdot \|a(x)\|), \tag{1.5}$$

then [under suitable smoothness assumptions on $a(x)$ and $b(x)$] the conclusion of Theorem 1.1 is still correct.

Note that (1.5) is close to $|b(x)| \cdot |x| = o(\|a(x)\|)$. Though our methods are confined to dimension 2, there is little doubt that analogous results hold in higher dimensions. Note also that comparison methods are not suited for the treatment of this conjecture. The reason is that in (1.3) as well as in (1.4), $v(x)$ does not depend monotonically on $b(x)$, and solutions of $Lv = -b$ cannot be bounded from above or below by enlarging or diminishing $b(x)$.

2. On random perturbations of dynamical systems. Consider a dynamical system $x = (x_t)$ in $\mathbb{R}^d$, given by the ordinary differential equation

$$dx_t = b(x_t)\,dt, \qquad t \geq 0. \tag{2.1}$$

We regard now our diffusion

$$dX_t = b(X_t)\,dt + \sigma(X_t)\cdot dW_t \tag{2.2}$$

as a random perturbation of (2.1), and the question arises as to whether the noisy component has a substantial effect on its long term behaviour. Special properties of such a random system, such as recurrence or transience, existence of a stationary distribution and so forth, are in general hardly accessible. It is therefore our intention to try an answer without referring to any particular property of (2.1) or (2.2), and rather look for an argument which tells us something like: "Whatever the behaviour of our diffusion may be, you cannot grasp it by neglecting the noisy part of the system and just analyzing the deterministic component." (This is in a different spirit than the celebrated theory of Ventcel and Freidlin on small random perturbations, which can be found in [24]. No large deviations will show up.)

This is not as hopeless as it might look. The idea is to look for new coordinates which are more useful for the discrimination between (2.1) and (2.2) than those given and which make it more or less evident that, in the long run, (x_t) and (X_t) behave differently.

Harmonic coordinates $u\colon \mathbb{R}^d \to \mathbb{R}^d$ are very useful for this purpose. Then, by means of the chain rule, (2.1) transforms into

$$du(x_t) = Du(x_t)\cdot b(x_t)\,dt, \tag{2.3}$$

whereas in view of $Lu \equiv 0$ and Itô's formula, (2.2) converts into

$$du(X_t) = Du(X_t)\cdot \sigma(X_t)\cdot dW_t. \tag{2.4}$$

Now evidently these two equations no longer exhibit any connection. In (2.4) there is no deterministic component present such that $u(X_t)$ is a local martingale. (2.3) on the contrary contains no noisy component such that one clearly cannot expect any relation between the long term behaviour of (2.3) and (2.4), whatever this behaviour may be. This argument becomes still more convincing if

$$Du(x) \to \mathrm{Id}$$

as $|x| \to \infty$. Then in view of (2.1)–(2.4) our change of coordinates has only a local effect on the systems up to the fact that in (2.4) the drift has been completely removed.

However, it is not $u(x)$ and $u(X)$ that we are aiming at, and certain aspects of (2.2) as well as (2.1) are not invariant under arbitrary changes of the coordinates. We have to take care that our consideration carries over to x and X. For this reason we restrict ourselves to coordinate transformations u such that

$$|u(x) - x| = o(|x|). \tag{2.5}$$

Loosely speaking we want $x_t \approx u(x_t)$, as well as $X_t \approx u(X_t)$, and we allow no global change of the state space's geometry. A (trivial) illustration is provided by the property that (X_t) is diffusive. This property, which means that asymptotically the mean square displacement of X_t is linear in t, is not coordinate-free; however, it is invariant under (2.5).

Thus Theorem 1.1 offers just the desired conclusion, and provides us with a tailor-made transformation of the state space. Condition (1.1) covers quite a broad range, such that our approach via harmonic coordinates is not as limited as one might think at first sight. The next example shows that on the borderline, where $|b(x)| \cdot |x|$ and $\|a(x)\|$ are of equal order, the picture changes. This example comes essentially from the theory of linear SDE, as developed by Has'minskii (cf. [12]) and others. (Compare also [16], where we explain in more detail, that systems with $|b(x)| \cdot |x| \asymp \|a(x)\|$ are just located between systems with dominating noise and systems where the noise is of secondary influence.)

EXAMPLE: NONEXISTENCE OF HARMONIC COORDINATES. Let $d = 2$ and let (2.2) be such that

$$dX = \beta X\,dt + \sigma_1 X\,dW_1 + \begin{pmatrix} 0 & -\sigma_2 \\ \sigma_2 & 0 \end{pmatrix} \cdot X\,dW_2$$

for $|X| \geq 1$ (to avoid a singularity at zero), with real numbers β, σ_1 and σ_2 and a two-dimensional standard Brownian motion $W = (W_1, W_2)^*$. Thus, letting $x = (x_1, x_2)^*$,

$$\sigma(x) = \begin{pmatrix} \sigma_1 x_1 & -\sigma_2 x_2 \\ \sigma_1 x_2 & \sigma_2 x_1 \end{pmatrix} = p(x) \cdot \begin{pmatrix} \sigma_1 |x| & 0 \\ 0 & \sigma_2 |x| \end{pmatrix}, \qquad |x| \geq 1,$$

where

$$p(x) = |x|^{-1} \begin{pmatrix} x_1 & -x_2 \\ x_2 & x_1 \end{pmatrix}$$

is the orthogonal matrix which rotates the vector $(1,0)^*$ into $x/|x|$. Then $a(x)$ has the eigenvalues $\sigma_1^2 |x|^2$ and $\sigma_2^2 |x|^2$ such that $|b(x)| \cdot |x| = \beta |x|^2$ and $\|a(x)\|$ are of the same order, namely, $|x|^2$. It is straightforward from Itô's formula that

$$d \log |X| = \left(\beta + \frac{\sigma_2^2 - \sigma_1^2}{2} \right) dt + \sigma_1\, dW_1$$

for $|X| \geq 1$. As a consequence, X is transient if $2\beta + \sigma_2^2 - \sigma_1^2 > 0$.

Next suppose that there are harmonic coordinates $u(x)$ such that $|u(x) - x| = o(|x|)$ and $Du(x) \to \mathrm{Id}$. It is no loss to assume $u(0) = 0$. Then $u(x)$ is uniquely determined; see Proposition 7.1 herein. Next the above equation for X has certain invariance properties for $|X| \geq 1$: it is rotationally invariant; furthermore, passing from $X = (X_1, X_2)^*$ to $\overline{X} = (-X_1, X_2)^*$, $\overline{X}$ obeys the above equation too (with W_2 replaced by $-W_2$). We will assume this also for $|X| \leq 1$. Then, due to uniqueness, $u(x)$ inherits all these symmetry properties, which imply that, in fact,

$$u(x) = f(|x|) \cdot x,$$

with $f(r) \to 1$ as $r \to \infty$ and, consequently,

$$Du(x) = p(x) \cdot \begin{pmatrix} \tau_1(|x|) & 0 \\ 0 & \tau_2(|x|) \end{pmatrix} \cdot p(x)^*,$$

with $\tau_1(r) = rf'(r) + f(r)$ and $\tau_2(r) = f(r)$. Since $Du(x) \to \mathrm{Id}$, $\tau_1(x) \to 1$ and $\tau_2(r) \to 1$ as $r \to \infty$. From $Lu \equiv 0$ and Itô's formula,

$$\begin{aligned} du(X) &= Du(X) \cdot \sigma(X) \cdot dW \\ &= \begin{pmatrix} \sigma_1 \tau_1(|x|)\, x_1 & -\sigma_2 \tau_2(|x|)\, x_2 \\ \sigma_1 \tau_1(|x|)\, x_2 & \sigma_2 \tau_2(|x|)\, x_1 \end{pmatrix} \cdot dW. \end{aligned}$$

As above, then

$$d \log|u(X)| = \frac{\sigma_2^2 \tau_2(|X|)^2 - \sigma_1^2 \tau_1(|X|)^2}{2}\, dt + \sigma_1 \tau_1(|X|)\, dW_1$$

for $|X| \geq 1$. Clearly $u(X)$ is recurrent for $\sigma_2^2 - \sigma_1^2 < 0$ since $\tau_1(r)$ and $\tau_2(r) \to 1$. Thus we may choose σ_1, σ_2 and β such that X is transient and $u(X)$ is recurrent at the same time. This being impossible, we conclude that such a transformation $u(x)$ does not exist.

To finish let us recall some related results that all support and illustrate our assertion that a small drift component [in the sense of (1.1) or (1.5)] has negligible influence on the random systems long term behaviour.

Asymptotic normality. Friedman derived in his monograph on stochastic differential equations [8] the following result:

Suppose that for $|x| \to \infty$,

$$a(x) \to \mathrm{Id}, \qquad b(x) = o(|x|^{-1}).$$

Then $X_t/\sqrt{t}$ *is asymptotically standard normal.*

Note that the drift component does not appear in the limit.

Invariant σ-fields. The following result is due to Gilbarg and Serrin [9] and Pinsky [19]:

(X_t) *has a trivial invariant σ-field if Assumption E holds and if for* $|x| \to \infty$,

$$|b(x)| \cdot |x| = O(\|a(x)\|).$$

In other words, every bounded harmonic function is constant. Pinsky additionally proves that the Martin boundary then is a singleton. To change this, a stronger drift has to be incorporated into the system.

Winding numbers. A celebrated theorem of Spitzer states that for a two-dimensional Brownian motion the winding number θ_t around zero up to time t is asymptotically Cauchy:

$$\frac{2\theta_t}{\log t} \to_d C$$

as $t \to \infty$, where C is standard Cauchy. LeGall and Yor [18] show that this result remains completely valid for the diffusion $dX_t = b(X_t)\,dt + dW_t$ in the plane provided that (1.5) holds.

3. The construction of harmonic coordinates. How can one obtain harmonic coordinates? As explained above we transform the equation $Lu \equiv 0$ by passing over to $v(x) = u(x) - x$, respectively, to the components of v. We are thus confronted with the following problem:

PROBLEM. Let $g\colon \mathbb{R}^d \to \mathbb{R}$ be continuous. Find conditions on g such that the equation

$$Lf = -g$$

has a solution $f\colon \mathbb{R}^d \to \mathbb{R}$ such that for $|x| \to \infty$,

$$f(x) = o(|x|).$$

[I find it helpful to leave out of account for the moment that $g(x)$ was obtained from the vector field $b(x)$.] For convenience we shall assume in the rest of the paper

$$\|a(x)\| = 1$$

for all x. Actually this normalization does not change the problem.

Our treatment of this problem is based on a probabilistic construction. It starts from the powerful existence and uniqueness results due to Stroock and Varadhan (see Theorem 5 (24.1) in [21] for instance, or consult [23]). They tell us that (2.2) has a unique weak solution (X_t) which is a strong Markov process. Let

$$\tau(M) = \inf\{t \geq 0\colon |X_t| = M\}.$$

Fix real numbers $0 < r_1 < r_2 < \cdots$ such that $r_n \uparrow \infty$ and consider the stopping times

$$\tau_j = \tau(r_j), \qquad j = 1, 2, \dots .$$

Then

$$f_m(x) = E_x \int_0^{\tau_m} g(X_t)\,dt - E_0 \int_0^{\tau_m} g(X_t)\,dt \tag{3.1}$$

(as usual, the index at the expectation denotes the starting position, respectively, the distribution of X_0) fulfills the equation

$$Lf_m(x) = -g(x) \quad \text{for } |x| < r_m.$$

We shall take away equal parts from both expectations in (3.1) [which is inspired by the consideration leading to (1.3)] and then let $m \to \infty$. Fix $k \in \mathbb{N}$

and $x \in \mathbb{R}^d$ such that $|x| < r_k$. Then by means of the strong Markov property,

$$
\begin{aligned}
f_m(x) &= (E_x - E_0)\int_0^{\tau_k} g(X_t)\,dt + \sum_{j=k}^{m-1} (E_x - E_0)\int_{\tau_j}^{\tau_{j+1}} g(X_t)\,dt \\
&= (E_x - E_0)\int_0^{\tau_k} g(X_t)\,dt + \sum_{j=k}^{m-1} \left(E_{\mu_j} - E_{\nu_j}\right)\int_0^{\tau_{j+1}} g(X_t)\,dt,
\end{aligned}
\tag{3.2}
$$

where μ_j and ν_j are the distributions of X_{τ_j}, given $X_0 = x$ and $X_0 = 0$, respectively. (Note that due to the shift of time, τ_j is replaced by 0.) Further let

$$
V(\mu, \nu) = \frac{1}{2}\int \left| \frac{d\mu}{d(\mu+\nu)} - \frac{d\nu}{d(\mu+\nu)} \right| d(\mu+\nu)
$$

be the distance of probability measures μ and ν in total variation and introduce the probability measures

$$
\alpha_j = \frac{\mu_j - \mu_j \wedge \nu_j}{V(\mu_j, \nu_j)}, \qquad \beta_j = \frac{\nu_j - \mu_j \wedge \nu_j}{V(\mu_j, \nu_j)}.
$$

Then, if we remove from the last two expectations in (3.2) the quantity $E_{\mu_j \wedge \nu_j} \int_0^{\tau_{j+1}} g(X_t)\,dt$, we obtain

$$
\begin{aligned}
f_m(x) &= (E_x - E_0)\int_0^{\tau_k} g(X_t)\,dt \\
&\quad + \sum_{j=k}^{m-1} V(\mu_j, \nu_j)\left(E_{\alpha_j} - E_{\beta_j}\right)\int_0^{\tau_{j+1}} g(X_t)\,dt.
\end{aligned}
\tag{3.3}
$$

α_j and β_j are probability measures with support on the sphere around zero with radius r_j. Note that μ_j, α_j and β_j depend on x. The expectations in (3.3) can be estimated rather precisely by means of the following lemma:

LEMMA 3.1. *Let $d \geq 2$ and let Assumption E be satisfied. Furthermore, assume*

$$
|x| \cdot |b(x)| = o(1)
$$

for $|x|$ large. Then there is a $\delta > 0$ such that for $\varepsilon < \delta$, $m \in \mathbb{N}$,

$$
\sup_{|x| \leq M} E_x\left[\int_0^{\tau(M)} |X_t|^{-1-\varepsilon}\,dt\right]^m = O(M^{m(1-\varepsilon)})
$$

as $M \to \infty$.

PROOF. Let $R_t = |X_t|$ and for $x \neq 0$,

$$
\lambda(x) = \frac{x^*}{|x|} \cdot a(x) \cdot \frac{x}{|x|}.
$$

By means of Itô's formula,

$$
dR^2 = (2X^* \cdot b(X) + \operatorname{trace} a(X))\,dt + 2R\lambda(X)^{1/2}\,dB
$$

with a suitable Brownian motion $B = (B_t)$. We first make use of a time change argument. Define κ_s, $s \ge 0$, by

$$s = \int_0^{\kappa_s} \lambda(X_t)\,dt.$$

Then for $\overline{X}_s = X_{\kappa_s}$, $\overline{R}_s = R_{\kappa_s} = |\overline{X}_s|$ and

$$\bar{\tau}(M) = \inf\{s \ge 0: \overline{R}_s = M\},$$

we obtain

$$E_x \int_0^{\tau(M)} R_t^{-1-\varepsilon}\,dt = E_x \int_0^{\bar{\tau}(M)} \overline{R}_s^{-1-\varepsilon} \frac{ds}{\lambda(\overline{X}_s)}.$$

Since $\lambda_{\min}(x) \le \lambda(x) \le \lambda_{\max}(x)$, we have in view of Assumption E, $1 = \|a(x)\| \le c\lambda(x)$ for all x and a suitable $c > 0$. Therefore,

$$E_x\left[\int_0^{\tau(M)} R_t^{-1-\varepsilon}\,dt\right]^m \le cE_x\left[\int_0^{\bar{\tau}(M)} \overline{R}_s^{-1-\varepsilon}\,ds\right]^m. \tag{3.4}$$

Note furthermore that

$$d\overline{R}^2 = \left(2\overline{X}^* \cdot b(\overline{X}) + \operatorname{trace} a(\overline{X})\right)\frac{ds}{\lambda(\overline{X})} + 2\overline{R}\,d\overline{B}$$

for some Brownian motion $\overline{B}$. Now, since $d \ge 2$, in view of Assumption E there is a $0 < \delta < 1$ such that

$$\begin{aligned} \operatorname{trace} a(x) &\ge \lambda_{\max}(x) + \lambda_{\min}(x) \\ &\ge (1 + 2\delta)\lambda_{\max}(x) \ge (1 + 2\delta)\lambda(x). \end{aligned}$$

Also,

$$x^* \cdot b(x) = o(1) = o(\lambda(x))$$

for $|x| \to \infty$ as well as $|x| \to 0$. Thus there are numbers $0 < \alpha < \beta$ and a differentiable function $\delta(z)$, $z \ge 0$, such that

$$2x^* \cdot b(x) + \operatorname{trace} a(x) \ge (1 + \delta(|x|))\lambda(x) \quad \text{for all } x$$

and

$$\delta(z) = \delta \quad \text{for } z < \alpha \text{ or } z > \beta.$$

Let us define a diffusion (Z_s) on $\mathbb{R}^+$ by means of the equations $Z_0 = \overline{R}_0$ and

$$dZ^2 = (1 + \delta(Z))\,ds + 2Z\,d\overline{B}.$$

In the case $\delta(z) = \delta$ for all $z \ge 0$, this equation is known to have a unique solution (a Bessel process). This fact is nontrivial, since the equation is degenerate at zero. The proof translates literally to our situation, since $\delta(z) = \delta$ for small z (see [21], Section V.48 for details). Using a suitable comparison theorem (as that of Ikeda and Watanabe; cf. [21], Theorem V.43), we conclude that $Z_s \le \overline{R}_s$ for all $s \ge 0$. Therefore,

$$E_x\left[\int_0^{\tau(M)} \overline{R}_t^{-1-\varepsilon}\, ds\right]^m \le E_z\left[\int_0^{\eta(M)} Z_s^{-1-\varepsilon}\, ds\right]^m \tag{3.5}$$

if $|x| = z < M$, with

$$\eta(M) = \inf\{s \ge 0: Z_s = M\}.$$

Now let $N > 0$ and consider the function

$$f(z) = E_z \int_0^{\eta(M)} N \wedge Z_s^{-1-\varepsilon}\, ds, \qquad z \ge 0.$$

Since $dZ = (\delta(Z)/2Z)\,ds + d\overline{B}$ for $Z \neq 0$, f satisfies on $(0, M)$ the equation

$$\frac{1}{2} f''(z) + \delta(z)\frac{f'(z)}{2z} = -N \wedge z^{-1-\varepsilon}.$$

Furthermore, since $0 < \delta < 1$, zero is an instantaneously reflecting boundary for (Z_s) (i.e., $\{s > 0\colon Z_s = 0\}$ has a.s. Lebesgue measure 0. For the Bessel process this is proved in [21], Section V.48). Thus

$$f'(0) = 0, \qquad f(M) = 0.$$

Therefore, f is uniquely determined and is given by

$$\begin{aligned}
f(z) &= 2\int_z^M du \int_0^u dv\, N \wedge v^{-1-\varepsilon} \exp\left(-\int_v^u \delta(w)\frac{dw}{w}\right) \\
&\le 2\int_0^M du \int_0^u dv\, v^{-1-\varepsilon} \exp\left(-\int_v^u (\delta - \gamma I(\alpha \le w \le \beta))\frac{dw}{w}\right) \\
&\le 2\left(\frac{\beta}{\alpha}\right)^{\gamma} \int_0^M du\, u^{-\delta} \int_0^u v^{-1-\varepsilon+\delta}\, dv
\end{aligned}$$

if γ is sufficiently large. Consequently, if $\varepsilon < \delta$,

$$\sup_{z \le M} E_z \int_0^{\eta(M)} Z_s^{-1-\varepsilon}\, ds \le \frac{2(\beta/\alpha)^{\gamma}}{(\delta - \varepsilon)(1 - \varepsilon)} M^{1-\varepsilon}.$$

Combining this estimate with (3.4) and (3.5) the desired result follows for $m = 1$. For $m > 1$ we proceed by induction. Let $h(x) = |x|^{-1-\varepsilon}$. Then

$$\begin{aligned}
&E_x\left[\int_0^{\tau(M)} h(X_t)\, dt\right]^m \\
&\quad = mE_x \int_0^{\tau(M)} h(X_t)\left(\int_t^{\tau(M)} h(X_s)\, ds\right)^{m-1} dt \\
&\quad = m\int_0^{\infty} E_x\left\{h(X_t)\cdot I(t < \tau(M)) E_{X_t}\left[\int_0^{\tau(M)} h(X_s)\, ds\right]^{m-1}\right\} dt \\
&\quad \le \text{const}\, M^{(1-\varepsilon)(m-1)} E_x \int_0^{\tau(M)} h(X_t)\, dt \\
&\quad = O(M^{(1-\varepsilon)m})
\end{aligned}$$

uniformly for $|x| \le M$, and our claim follows. □

NOTE. In a similar manner one can obtain uniform bounds on

$$E_x \exp\left(\lambda M^{\varepsilon-1} \int_0^{\tau(M)} |X_t|^{-1-\varepsilon}\, dt\right)$$

for small $\lambda > 0$.

The other thing to do is to estimate $V(\mu_j, \nu_j)$ in order to let $m \to \infty$ in (3.3). This will be done in the next section, but first we would like to indicate what kind of estimate will do for our purposes. Let $M > 0$ and consider $y, z \in \mathbb{R}^d$ such that $|y|, |z| < M$. Let further $\mu_{y,M}$ and $\mu_{z,M}$ be the distribution of $X_{\tau(M)}$, given $X_0 = y$ and $X_0 = z$, respectively, with distance in total variation

$$V_M(y, z) = V(\mu_{y,M}, \mu_{z,M}).$$

Note that for $|y|, |z| < N < M$,

$$V_M(y, z) \le V_N(y, z) \cdot \sup_{|u|=|v|=N} V_M(u, v). \tag{3.6}$$

For the proof use

$$\begin{aligned} V_M(y, z) &= \sup_B |\mu_{y,M}(B) - \mu_{z,M}(B)| \\ &= \sup_B \left| \int \mu_{u,M}(B) \mu_{y,N}(du) - \int \mu_{u,M}(B) \mu_{z,N}(du) \right|. \end{aligned}$$

Taking away $\int \mu_{u,M}(B)(\mu_{y,N} \wedge \mu_{z,N})(du)$, we get the desired estimate

$$\begin{aligned} V_M(y, z) &\le \sup_B \left(\sup_{|u|=|v|=N} |\mu_{u,M}(B) - \mu_{v,M}(B)| \cdot V_N(y, z) \right) \\ &\le \sup_{|u|=|v|=N} V_M(u, v) \cdot V_N(y, z). \end{aligned}$$

PROPOSITION 3.2. *Let $d \ge 2$, let Assumption E be satisfied and suppose*:

(i) $|b(x)| = o(|x|^{-1})$.

(ii) *There are numbers $\varepsilon > 0$, $c > 0$, such that for all x*,

$$|g(x)| \le c(1 + |x|)^{-1-\varepsilon}.$$

(iii) *For any $\alpha > 0$ there are numbers $r > 1$, $\lambda > 0$, such that*

$$V_{rM}(y, z) \le r^{\alpha-1}$$

if only $\lambda \le |y|, |z| \le M$.

Then, as $m \to \infty$, f_m converges pointwise to a limiting function f. f is a C^2-function

$$Lf = -g$$

on $\mathbb{R}^d$ and as $|x| \to \infty$,

$$f(x) = o(|x|).$$

PROOF. Let $\alpha > 0$ and fix $r > 1$, $\lambda > 0$ according to assumption (iii). It is convenient to choose $r_j = r^j$ now. Fix $k \in \mathbb{N}$ such that $r^k > \lambda$ and let $|x| < r^k$. Then in view of (3.6) and assumption (iii),

$$V(\mu_j, \nu_j) \le r^{(\alpha-1)(j-k)}$$

if $j \ge k$. Therefore, in view of (3.3) and Lemma 3.1, if $k < m \le n$,

$$|f_n(x) - f_m(x)| \le C \sum_{j=m}^{n-1} r^{(\alpha-1)(j-k)} \cdot r^{(1-\varepsilon)(j+1)}$$

for some $C > 0$. By assumption we may choose α arbitrarily small, say $\alpha = \varepsilon/2$. Then $|f_n(x) - f_m(x)| \to 0$ as $m, n \to \infty$ and $f_m(x)$ is convergent to the limiting function

$$f(x) = (E_x - E_0)\int_0^{\tau_k} g(X_t)\,dt + \sum_{j=k}^{\infty} V(\mu_j, \nu_j)\big(E_{\alpha_j} - E_{\beta_j}\big)\int_0^{\tau_{j+1}} g(X_t)\,dt.$$

Since all estimates are uniform for $|x| < r^k$, we have uniform convergence on compact sets. As is well known, $Lf_m = -g$ on $|x| < r^m$ then entails $f \in C^2$ and

$$Lf = -g$$

in the limit. Furthermore, using Lemma 3.1 again and now choosing k such that $r^{k-1} \le |x| < r^k$,

$$\begin{aligned} |f(x)| &\le \left|(E_x - E_0)\int_0^{\tau_k} g(X_t)\,dt\right| + C\sum_{j=k}^{\infty} r^{(\alpha-1)(j-k)} r^{(1-\varepsilon)(j+1)} \\ &= O(r^{k(1-\varepsilon)} + r^{k(1-\alpha)}) \\ &= O\big(|x|^{1-\varepsilon} + |x|^{1-\alpha}\big) \end{aligned}$$

and the desired result follows. □

4. The coupling. Let again $M > 0$ and consider points $y, z \in \mathbb{R}^2$ such that $|y|, |z| < M$. In order to apply Proposition 3.2, it is the main task to estimate the variational distance $V_M(y, z)$. We do this via a coupling: Let X and X' be two versions of our diffusion, with $X_0 = y$ and $X'_0 = z$. Let $\tau(M)$ be as above and let

$$\tau'(M) = \inf\{t \ge 0: |X'_t| = M\}$$

be the corresponding hitting time for X'. Then we have the usual estimate

$$V_M(y, z) \le \Pr_{y,z}\big(X_t \ne X'_t \text{ for all } t \le \tau(M) \wedge \tau'(M)\big).$$

The argument is: If X_t and X'_t meet at some instance $t \le \tau(M) \wedge \tau'(M)$, then we may couple X'_t to X_t from this instance. These paths then give a common contribution to the distributions of $X_{\tau(M)}$ and $X'_{\tau'(M)}$. Thus one would like to construct X and X' simultaneously in such a way that the above probability is sufficiently small. Additional flexibility is given by the observation that instead of X and X', we may as well consider random time changes of X and X'. They do not change $X_{\tau(M)}$ (resp. $X'_{\tau'(M)}$) and thus have no effect on $V_M(y, z)$.

Before going into details, let us try to explain our choice of the time changes in a heuristic manner. Recall first that for a standard Brownian motion W_t in $\mathbb{R}^d$ the norm $R_t = |W_t|$ constitutes a Bessel process. In dimension greater than 1 this is not a martingale, but dR_t contains a drift component $cR_t^{-1}\,dt$, with $c = (d-1)/2$. This makes it harder for R_t to reach zero (in the case $c \geq 2$, even impossible; cf. [21], Chapter IV.35). For $|X_t - X'_t|$ we must be prepared for a similar effect—a singularity of the drift at $X - X' = 0$—which may prevent X and X' from meeting. However, this effect vanishes if those local random oscillations of X_t and X'_t perpendicular to $X_t - X'_t$ coincide. For example, let $X = (W, B)^*$ and $X' = (W', B)^*$, with independent standard Brownian motions W, W' and B in $\mathbb{R}$ and $W_0 \neq W'_0$. Then $|X_t - X'_t| = |W_t - W'_t|$ contains no singular drift at zero. (Here the "perpendicular" noise is B.) Now in dimension 2 one can synchronize this "perpendicular" noise by means of a suitable time change. This will be explained now in detail. The key formula which tells us that no singular drift appears is contained in the proof of Lemma 4.1 below. [It is $(dY - dZ)^* \cdot P \cdot (dY - dZ) = 0$, where Y and Z will be obtained form X and X' by certain time changes and P denotes the projection matrix onto the vector perpendicular to $Y - Z$.]

Thus let $d = 2$ in the sequel. We introduce some notation. For $z = (z_1, z_2)^* \neq 0$,

$$p(z) = |z|^{-1}\begin{pmatrix} z_1 & -z_2 \\ z_2 & z_1 \end{pmatrix}$$

denotes the orthogonal matrix which rotates the vector $(1,0)^*$ into $z/|z|$. For $x, z \in \mathbb{R}^2$, let

$$a(x,z) = \big(a_{ij}(x,z)\big) = p(z)^* \cdot a(x) \cdot p(z)$$

and

$$\sigma(x,z) = \begin{pmatrix} (\det a(x,z))^{1/2} a_{22}^{-1/2}(x,z) & a_{12}(x,z)a_{22}^{-1/2}(x,z) \\ 0 & a_{22}^{1/2}(x,z) \end{pmatrix}.$$

$a(x,z)$ is the diffusion matrix of system (2.2) at point x with respect to the basis $z/|z|$ and $z^{\perp}/|z|$, and

$$\sigma(x,z) \cdot \sigma(x,z)^* = a(x,z).$$

Now to the coupling. Let B, B', B'' be independent Brownian motions in $\mathbb{R}$. B will be used to generate the "perpendicular" random noise. Consider a continuous process (Y, Z) in $\mathbb{R}^2 \times \mathbb{R}^2$ with $(Y_0, Z_0) = (y, z)$ which satisfies the stochastic equations

$$
\begin{aligned}
dY &= \frac{b(Y)}{a_{22}(Y, Y-Z)}\,dt + a_{22}^{-1/2}(Y, Y-Z)\,p(Y-Z) \\
&\quad \cdot \sigma(Y, Y-Z) \cdot \begin{pmatrix} dB' \\ dB \end{pmatrix}, \\
dZ &= \frac{b(Z)}{a_{22}(Z, Y-Z)}\,dt + a_{22}^{-1/2}(Z, Y-Z)\,p(Y-Z) \\
&\quad \cdot \sigma(Z, Y-Z) \cdot \begin{pmatrix} dB'' \\ dB \end{pmatrix}
\end{aligned}
\tag{4.1}
$$

in the period of time from 0 to $\inf\{t\colon Y_t = Z_t\}$ before Y and Z meet for the first time. Afterward we require

$$
dY = b(Y)\,dt + \sigma(Y) \cdot \begin{pmatrix} dB' \\ dB \end{pmatrix} \tag{4.2}
$$

and $Z = Y$ such that Y and Z stay together after they have met. The question arises whether such a process exists at all. It turns out that this can be achieved on a suitable probability space. It seems that this is not covered by known results [note that due to coupling, (4.1) is degenerate; furthermore, the coefficients fail to be continuous at the moment when Y and Z meet]. Nevertheless, some routine considerations from the "martingale problem" point of view will settle the problem. In order not to interrupt our line of argument, we postpone this question and come back to it in the Appendix.

The question as to whether (4.1) and (4.2) can be solved uniquely and whether (Y, Z) constitutes a diffusion process is more subtle. However, this is of no importance for us: we shall make use of (4.1) and (4.2) solely. This is the point: If we change time, before Y and Z meet, $X_t = Y_{\kappa_t}$, $U_t = Z_{\kappa_t}$ with

$$
t = \int_0^{\kappa_t} a_{22}^{-1}(Y_s, Y_s - Z_s)\,ds, \tag{4.3}
$$

then from (4.1),

$$
dX = b(X)\,dt + p(X-U) \cdot \sigma(X, X-U) \cdot \begin{pmatrix} d\overline{B}' \\ d\overline{B} \end{pmatrix}
$$

with independent Brownian motions $\overline{B}, \overline{B}'$. This gives the desired diffusion. Since

$$
p(X_t - U_t) \cdot \sigma(X_t, X_t - U_t) \cdot \sigma^*(X_t, X_t - U_t) \cdot p^*(X_t - U_t) = a(X_t),
$$

by means of Itô's formula,

$$
f(X_t) - f(X_0) - \int_0^t Lf(X_s)\,ds
$$

is a martingale for every C^∞-function f with compact support. In other words, X solves the martingale problem for L (see [21], Chapter V.19). Now under Assumptions A and E, and since $\|a(x)\|$ is assumed to be 1 everywhere, we may apply the fundamental uniqueness result of Stroock and Varadhan (see [23] or Theorem V (24.1) in [21]), which tells us that (X_t) in fact is the unique diffusion process belonging to the operator L. Similarly, Z_t can be viewed as our diffusion—changed in time—and our construction offers the desired coupling.

Let us consider now

$$D_t = |Y_t - Z_t|.$$

LEMMA 4.1. *Let $d = 2$ and let Assumption E be satisfied. Then, as long as $D_t > 0$,*

$$dD = d(Y,Z)\,dt + s(Y,Z)\,dB_D.$$

Here B_D denotes a standard Brownian motion on $\mathbb{R}$, $s(y,z)$ is a function, which is bounded away from zero (and from above) on $\mathbb{R}^4$ and

$$|d(y,z)| \le c|b(y)| + c|b(z)|$$

for some $c > 0$ and all y, z.

PROOF. Itô's formula yields (as long as $D_t > 0$)

$$dD = \left(\frac{Y-Z}{D}\right)^* \cdot (dY - dZ) + \frac{1}{2D}(dY - dZ)^* \cdot P \cdot (dY - dZ).$$

$P = (\delta_{ij} - D^{-2}(Y_i - Z_i)(Y_j - Z_j))_{i,j=1,2}$ is the projection matrix onto the vector perpendicular to $(Y - Z)$ (δ_{ij} is Kronecker's symbol). Therefore,

$$p^*(Y-Z)\cdot P \cdot p(Y-Z) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Note also that the second component of the stochastic differential

$$\begin{aligned} dU = {} & a_{22}^{-1/2}(Y, Y-Z)\,\sigma(Y, Y-Z)\cdot \begin{pmatrix} dB' \\ dB \end{pmatrix} \\ & - a_{22}^{-1/2}(Z, Y-Z)\,\sigma(Z, Y-Z)\cdot \begin{pmatrix} dB'' \\ dB \end{pmatrix} \end{aligned}$$

vanishes. It follows that

$$(dY - dZ)^* \cdot P \cdot (dY - dZ) = (dU)^* \cdot \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \cdot dU = 0$$

such that the second term in dD vanishes. As to the first term, note that

$$p^*(Y-Z)\cdot \frac{Y-Z}{D} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

such that

$$\left(\frac{Y-Z}{D}\right)^* \cdot (dY - dZ)$$
$$= d(Y,Z)\,dt + (1,0)\cdot dU$$
$$= d(Y,Z)\,dt + \frac{\sqrt{\det a(Y,Y-Z)}}{a_{22}(Y,Y-Z)}\,dB' - \frac{\sqrt{\det a(Z,Y-Z)}}{a_{22}(Z,Y-Z)}\,dB''$$
$$+\left(\frac{a_{12}(Y,Y-Z)}{a_{22}(Y,Y-Z)} - \frac{a_{12}(Z,Y-Z)}{a_{22}(Z,Y-Z)}\right)dB,$$

where

$$d(Y,Z) = \left(\frac{b(Y)}{a_{22}(Y,Y-Z)} - \frac{b(Z)}{a_{22}(Z,Y-Z)}\right)^* \cdot \frac{Y-Z}{D}.$$

Thus the desired stochastic equation is obtained with

$$s^2(y,z) = \frac{\det a(y,y-z)}{a_{22}(y,y-z)^2} + \frac{\det a(z,y-z)}{a_{22}(z,y-z)^2}$$
$$+\left(\frac{a_{12}(y,y-z)}{a_{22}(y,y-z)} - \frac{a_{12}(z,y-z)}{a_{22}(z,y-z)}\right)^2.$$

To show that $s(y,z)$ is bounded away from zero, note that $a(y,y-z)$ inherits the property of Assumption E from $a(x)$. Therefore,

$$a_{22}(y,y-z) \le \lambda_{\max}(y).$$

Also

$$\det a(y,y-z) = \lambda_{\min}(y)\cdot\lambda_{\max}(y),$$

which in view of Assumption E leads to

$$0 < C^{-1} \le \frac{\lambda_{\min}(y)}{\lambda_{\max}(y)} \le s(y,z)^2.$$

The estimate on $d(y,z)$ is also a consequence of Assumption E. □

We now divide the estimation of $V_M(y,z)$ into several parts. Denote

$$\rho(M) = \inf\{t \ge 0 : |Y_t| = M\}, \qquad \rho'(M) = \inf\{t \ge 0 : |Z_t| = M\}.$$

LEMMA 4.2. *Let $d = 2$ and let Assumption E be satisfied. Furthermore, $|x|^{1+\varepsilon}\cdot|b(x)| = O(1)$ for some $\varepsilon > 0$. Then, for any $\delta < \varepsilon$ and $c > 0$,*

$$\Pr_{y,z}\left(\int_0^{\rho(rM)} |Y_t|^{-1-\varepsilon}\,dt \ge c(rM)^{1-\delta}\right) = o(r^{-1})$$

as $r \to \infty$ uniformly in $1 \le |y|, |z| \le M < \infty$.

PROOF. Let us again apply the time change $X_t = Y_{\kappa_t}$, with κ_t as in (4.3). Since $\|a(x,z)\| = \|a(x)\| = 1$,

$$\int_0^{\rho(rM)} |Y_t|^{-1-\varepsilon}\,dt = \int_0^{\tau(rM)} |X_t|^{-1-\varepsilon} a_{22}(X_t, X_t - U_t)\,dt$$
$$\le \int_0^{\tau(rM)} |X_t|^{-1-\varepsilon}\,dt.$$

The claim follows now from Lemma 3.1 and the Markov inequality. □

LEMMA 4.3. *Under the assumptions of Lemma* 4.2, *for any* $\delta > 0$,

$$\Pr_{y,z}\big(\rho(rM) \le r^{2-\delta}M^2\big) = o(r^{-1})$$

as $r \to \infty$, *uniformly in* $1 \le |y|, |z| \le M < \infty$.

PROOF. Let Y_{1t} and Y_{2t} be the components of Y_t. Then the event $\{\rho(rM) \le r^{2-\delta}M^2\}$ is (for r large enough) contained in

$$\begin{aligned}
&\left\{\int_0^{\rho(rM)} |b(Y_t)| a_{22}^{-1}(Y_t, Y_t - Z_t)\,dt \ge \frac{rM}{4}\right\}\\
&\quad\cup \bigcup_{i=1}^{2}\left\{|Y_{it} - Y_{i0}| \ge \frac{rM}{2} \text{ and } \int_0^t |b(Y_t)| a_{22}^{-1}(Y_t, Y_t - Z_t)\,dt\right.\\
&\qquad\qquad\qquad\qquad \left.\le \frac{rM}{4} \text{ for some } t \le r^{2-\delta}M^2\right\}\\
&\quad = A \cup B_1 \cup B_2 \quad (\text{say}).
\end{aligned}$$

Since $a_{22}(y, y-z) \ge c\|a(y)\| = c$ by uniform ellipticity, $|b(y)a_{22}^{-1}(y, y-z)| = O(|y|^{-1-\varepsilon})$ by assumption; hence, $\Pr_y(A) = o(r^{-1})$ in view of Lemma 4.2. As to B_1, we notice that

$$dY_1 = \frac{b_1(Y)}{a_{22}(Y, Y-Z)}\,dt + \frac{a_{11}(Y)^{1/2}}{a_{22}(Y, Y-Z)^{1/2}}\,d\overline{B}$$

for a suitable Brownian motion $\overline{B}$, as follows from (4.1). Thus

$$\Pr_{y,z}(B_1) \le \Pr_y\left(\left|\int_0^t \frac{a_{11}(Y)^{1/2}}{a_{22}(Y, Y-Z)^{1/2}}\,d\overline{B}\right| \ge \frac{rM}{4} \text{ for some } t \le r^{2-\delta}M^2\right).$$

In view of Assumption E, the corresponding quadratic-variation process

$$\int_0^t \frac{a_{11}(Y)}{a_{22}(Y, Y-Z)}\,dt$$

can be estimated from above by ct for all $t \geq 0$ and a suitable $c > 0$, such that in view of a martingale estimate (Theorem IV (37.8) in [21]),

$$\Pr_{y,z}(B_1) \leq 2\exp(-\text{const}\, r^{\delta}) = o(r^{-1})$$

uniformly in $1 \leq |y| \leq M < \infty$. $\Pr_y(B_2)$ is estimated similarly. □

LEMMA 4.4. *Under the assumptions of Lemma* 4.2, *for any* $\delta < \varepsilon$,

$$\Pr_{y,z}\big(D_t > 0 \text{ for all } t \leq r^{2-\delta}M^2\big) = O(r^{\delta/2-1})$$

for $r \to \infty$, *uniformly in* $r^{1/\delta-1} \leq |y|, |z| \leq M < \infty$.

PROOF. The event $\{D_t > 0 \text{ for all } t \leq r^{2-\delta}M^2\}$ is contained in

$$\begin{aligned}
&\left\{\int_0^{\rho(rM)} |Y_t|^{-1-\varepsilon}\,dt + \int_0^{\rho'(rM)} |Z_t|^{-1-\varepsilon}\,dt \geq cM\right\}\\
&\qquad \cup \big\{\rho(rM) \leq r^{2-\delta}M^2\big\} \cup \big\{\rho'(rM) \leq r^{2-\delta}M^2\big\}\\
&\qquad \cup \left\{D_t > 0 \text{ for all } t \leq r^{2-\delta}M^2,\ \int_0^{r^{2-\delta}M^2} |d(Y_t, Z_t)|\,dt \leq M\right\}\\
&\qquad = A_0 \cup A_1 \cup A_2 \cup A_3 \quad \text{(say)}
\end{aligned}$$

for some $c > 0$, as an inspection of Lemma 4.1 shows.

As to A_0, note that $1 \leq r^{1/\delta-1} \leq M$ implies $M \geq (rM)^{1-\delta}$ and $\Pr_{y,z}(A_0) = o(r^{-1})$ follows from Lemma 4.2 and the analogous statement for (Z_t). $\Pr(A_1 \cup A_2) = o(r^{-1})$ is just Lemma 4.3 and the analogous statement for $\rho'(rM)$. Next note that $D_0 = |y - z| \leq 2M$. Thus, in view of Lemma 4.1,

$$\Pr_{y,z}(A_3) \leq \Pr_{y,z}\left(\int_0^t s(Y, Z)\,dB_D \geq -3M \text{ for all } t \leq r^{2-\delta}M^2\right).$$

However, $\int_0^t s(Y,Z)\,dB_D$ can be viewed as a Brownian motion at (random) time $\int_0^t s^2(Y,Z)\,dt$. Since s is bounded away from zero, we obtain

$$\Pr_{y,z}(A_3) \leq \Pr\big(B(t) \leq 3M \text{ for all } t \leq cr^{2-\delta}M^2\big)$$

for some $c > 0$ and a standard Brownian motion $B(t)$. The reflection principle then yields

$$\begin{aligned}
\Pr_{y,z}(A_3) &\leq \Pr\big(|B(cr^{2-\delta}M^2)| \leq 3M\big)\\
&= \Pr\big(|B(c)| \leq 3r^{\delta/2-1}\big) \leq 6c^{-1}r^{\delta/2-1}
\end{aligned}$$

and the claim follows. □

Putting together our considerations, we are now in the position to prove the existence part of our theorem:

Let $d = 2$, *let Assumptions A and E be satisfied and assume* $|b(x)| = O(|x|^{-1-\varepsilon})$ *for some* $\varepsilon > 0$. *Then there is a mapping* $u\colon \mathbb{R}^2 \to \mathbb{R}^2$ *such that* $Lu \equiv 0$ *and* $|u(x) - x| = o(|x|)$.

In fact, using Lemma 4.3 [together with the analogous statement for $\rho'(rM)$] and Lemma 4.4, for sufficiently small $\delta > 0$,

$$\begin{aligned} V_{rM}(y,z) &\le \Pr_{y,z}\big(D_t > 0 \text{ for all } t \le \rho(rM) \wedge \rho'(rM)\big) \\ &\le \Pr_{y,z}\big(D_t > 0 \text{ for all } t \le r^{2-\delta}M^2\big) \\ &\qquad + \Pr_{y,z}\big(\rho(rM) \le M^2 r^{2-\delta} \text{ or } \rho'(rM) \le M^2 r^{2-\delta}\big) \\ &= O(r^{\delta/2-1}) \end{aligned}$$

uniformly in $r^{1/\delta-1} \le |y|, |z| \le M$. This implies condition (iii) of Proposition 3.2, which thus may be applied. Hence there exists a mapping $v\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that (componentwise)

$$Lv = -b, \qquad |v(x)| = o(|x|).$$

Letting $u(x) = v(x) + x$, the desired mapping $u\colon \mathbb{R}^2 \to \mathbb{R}^2$ is obtained.

5. Estimates for the gradient. In this section we prove the following result.

PROPOSITION 5.1. *Let $d = 2$, let Assumptions A and E be satisfied, let $\|a(x)\| = 1$ and*

$$|x| \cdot |b(x)| = o(1).$$

If $f\colon \mathbb{R}^2 \to \mathbb{R}$ in a C^2-function such that for $|x| \to \infty$,

$$f(x) = o(|x|)$$

and

$$Lf(x) = o(|x|^{-1}),$$

then

$$|\mathrm{grad}\, f(x)| = o(1).$$

Of course the corresponding statement in our theorem follows. Such a result is usually called an a priori estimate. Classical a priori bounds as those of Schauder (compare the monograph [10]) are not fully adapted for our purposes. We therefore give our own proof. We approach it in a probabilistic manner, using once more our coupling. It was Cranston [5] who realized that one may obtain a priori estimates by means of couplings.

PROOF OF PROPOSITION 5.1. Denote $g = -Lf$. In order to estimate the difference $f(y) - f(z)$, we express it by our coupling (Y, Z). Let again $Y_0 = y$, $Z_0 = z$, $d = |y - z| \le |y|/4$ and

$$\rho = \inf\left\{t \ge 0\colon |Y_t - y| = \frac{|y|}{2} \text{ or } |Z_t - y| = \frac{|y|}{2} \text{ or } D_t = 0\right\}.$$

We begin with estimating the expectation of ρ. In view of Lemma 4.1 and Itô's formula,

$$d\big(D_t(|y| - D_t)\big) = d(\text{martingale}) + \big\{(|y| - 2D_t)d(Y_t, Z_t) - s^2(Y_t, Z_t)\big\}\,dt.$$

Let $t \le \rho$. Then $|Y_t| \ge |y|/2$, $|Z_t| \ge |y|/2$ and $D_t \le |y|$ such that in view of Lemma 4.1 and $|y| \cdot |b(y)| = o(1)$ the term $(|y| - 2D_t)d(Y_t, Z_t)$ can be made arbitrarily small if only $|y|$ is chosen large enough. On the other hand, $s^2(Y, Z)$ is bounded away from zero. Thus, applying optional sampling, for $M > 0$ and a suitable $c > 0$,

$$\begin{aligned} E_{y,z}\, \rho \wedge M &\le c E_{y,z} \int_0^{\rho\wedge M} \bigl(s^2(Y,Z) - (|y| - 2D)d(Y,Z)\bigr)\, dt \\ &= cD_0(|y| - D_0) - cE_{y,z}\bigl(D_{\rho\wedge M}(|y| - D_{\rho\wedge M})\bigr). \end{aligned}$$

Using $D_{\rho\wedge M} \le |y|$ and letting $M \uparrow \infty$ we end up with

$$E_{y,z}\, \rho \le c|y - z| \cdot |y| \tag{5.1}$$

if only $|y|$ is large enough.

Next, to obtain a representation for $f(y)$, we observe that in view of (4.1),

$$df(Y) = -\frac{g(Y)}{a_{22}(Y, Y - Z)}\, dt + d(\text{local martingale}).$$

Note that grad f is contained in the stochastic differential on the right-hand side. It is not obvious whether this differential is derived from a martingale, since so far we have no control on the gradient. For $f(Y_{t\wedge\rho})$ this doubt does not exist, since $Y_{t\wedge\rho}$ stays in a bounded region, where grad f is bounded, such that the corresponding differential stems from a martingale. Another application of optional stopping yields

$$f(y) = E_{y,z} f(Y_{\rho\wedge M}) + E_{y,z} \int_0^{\rho\wedge M} \frac{g(Y)}{a_{22}(Y, Y - Z)}\, dt.$$

Since $|Y_{\rho\wedge M}| \le 2|y|$ and $E_{y,z}\, \rho < \infty$ in view of (5.1), we may pass to the limit $M \to \infty$:

$$f(y) = E_{y,z} f(Y_\rho) + E_{y,z} \int_0^{\rho} \frac{g(Y)}{a_{22}(Y, Y - Z)}\, dt,$$

and further

$$\begin{aligned} f(y) - f(z) &= E_{y,z}\bigl(f(Y_\rho) - f(Z_\rho)\bigr) \\ &\quad + E_{y,z} \int_0^{\rho} \left(\frac{g(Y)}{a_{22}(Y, Y - Z)} - \frac{g(Z)}{a_{22}(Z, Y - Z)} \right) dt. \end{aligned}$$

Since $t \le \rho$ entails $|Y_t|, |Z_t| \ge |y|/2$, it follows by assumption that $f(Y_\rho) = o(|y|)$ and $g(Y_t)a_{22}^{-1}(Y_t, Y_t - Z_t) = o(|y|^{-1})$. Since further $f(Y_\rho) = f(Z_\rho)$ in case $D_\rho = 0$,

$$|f(y) - f(z)| = o(|y|) \cdot \Pr_{y,z}(D_\rho \ne 0) + o(|y|^{-1}) E_{y,z}\, \rho. \tag{5.2}$$

It remains to estimate this probability. Let K be the cone of all points (x, d) in $\mathbb{R}^2 \times \mathbb{R}$ such that

$$d \ge 0 \quad \text{and} \quad 4|x - y| \le |y| - cd$$

for some given $c > 0$. Also let τ be the first instance that (Y_t, D_t) hits the boundary of K. If $t \le \tau$, $|Y_t - y| \le \frac{1}{4}|y|$ and $|Z_t - y| \le |Y_t - y| + D_t \le (\frac{1}{4} + 1/c)|y| \le |y|/3$ for large c. Therefore, $\tau \le \rho$, $\{D_\rho \ne 0\} \subset \{D_\tau \ne 0\}$ and

$$\Pr_{y,z}(D_\rho \ne 0) \le \Pr_{y,z}(D_\tau \ne 0).$$

Next let

$$h(Y_t, D_t) = 1 + |y|^{-2}\Big[16|Y_t - y|^2 - (|y| - cD_t)^2\Big].$$

Note that $h(Y_\tau, D_\tau) \ge 0$; furthermore, $D_\tau \ne 0$ implies $h(Y_\tau, D_\tau) = 1$. Thus

$$\Pr_{y,z}(D_\tau \ne 0) \le E_{y,z} h(Y_\tau, D_\tau).$$

This expectation will now be estimated by the optional stopping theorem. We want $h(Y_t, D_t)$ to be a supermartingale, as long as (Y_t, D_t) stays in the cone K. This can be achieved if c is chosen large enough (uniformly for all y). In fact, from Lemma 4.1 and Itô's formula,

$$\begin{aligned} dh(Y, D) &= d(\text{martingale}) \\ &\quad + |y|^{-2}\left\{32\frac{b(Y)^* \cdot (Y - y)}{a_{22}(Y, Y - Z)} + 16\frac{\operatorname{trace} a(Y, Y - Z)}{a_{22}(Y, Y - Z)}\right. \\ &\qquad\qquad \left. + 2c(|y| - cD)d(Y, Z) - c^2 s^2(Y, Z)\right\} dt. \end{aligned}$$

Note that $(Y_t, D_t) \in K$ implies $|Y_t| \ge \frac{1}{2}|y|$ and $D_t \le |Y_t|/c$; thus, $|Z_t| \ge |Y_t| - D_t \ge \frac{1}{4}|y|$ if c is large enough. Therefore, in view of our assumptions and since $s^2(Y, Z)$ is bounded away from zero, the last term in the above drift component dominates if c is sufficiently large. If now $d = |y - z|$ is smaller than $|y|/c$, $(Y_0, D_0) \in K$ and it follows by optional stopping that

$$E_{y,z} h(Y_\tau, D_\tau) \le h(y, d).$$

Altogether we obtain

$$\begin{aligned} \Pr_{y,z}(D_\rho \ne 0) \le h(y, d) &= 1 - |y|^{-2}(|y| - c|y - z|)^2 \\ &\le 2c\frac{|y - z|}{|y|}. \end{aligned}$$

Combining this estimate with (5.1) and (5.2), we end up with

$$\frac{|f(y) - f(z)|}{|y - z|} = o(1)$$

as $|y| \to \infty$, if only $|y - z| \le |y|/c$, and our claim follows. □

6. Bijectivity.

The main result of this section is

PROPOSITION 6.1. *If u: $\mathbb{R}^2 \to \mathbb{R}^2$ is a C^2-mapping such that*

(6.1) $$Lu \equiv 0$$

and for $|x| \to \infty$,

(6.2) $$|u(x) - x| = o(|x|),$$

$$|Du(x) - \mathrm{Id}| = o(1), \tag{6.3}$$

then u is a diffeomorphism, that is, a regular bijection from $\mathbb{R}^2$ onto $\mathbb{R}^2$.

PROOF. (i) To obtain surjectivity, fix $y \in \mathbb{R}^2$ and consider the mapping

$$F(x) = x - u(x) + y.$$

In view of (6.2), F maps B_r, the ball of radius r around the origin, into itself if only r is large enough. Applying Brouwer's fixpoint theorem, there is a $x_0 \in B_r$ such that $F(x_0) = x_0$. Thus $u(x_0) = y$, and u is surjective.

(ii) Let us show now that the gradient of u_1 is nowhere vanishing in the plane, where u_1 is the first component of the given mapping $u = (u_1, u_2)$. Consider the vector field

$$v(x) = \operatorname{grad} u_1(x).$$

We shall utilize the topological concept of an index. Recall that the index of a closed curve with respect to a vector field $v(x)$ in the plane is the total amount of rotation of the vector $v(x)/|v(x)|$ if x runs around the curve. If x_0 is an isolated singularity [i.e., $v(x_0) = 0$ and $v(x) \neq 0$, $x \neq x_0$, in some neighbourhood of x_0], its index $\operatorname{ind}(x_0)$ is defined to be the index of a sufficiently small circle surrounding it. ([1] contains a nontechnical introduction to this subject.) Now this is the idea of the proof: In view of (6.3), $v(x) \to (1,0)^*$ as $|x| \to \infty$. Therefore, a circle around zero with sufficiently large radius r has index 0. The theorem of the index sum then implies that

$$\sum_{|y|<r} \operatorname{ind}(y) = 0$$

provided $v(x)$ has only isolated singularities inside this circle. (The sum is taken over singular points y.) The index of a saddle point is negative. On the other hand, the index of a maximum or minimum is 1. In view of (6.1) and the strong maximum principle (see [10]), u_1 has no maxima or minima at all. Thus, the above index sum can be zero only if there are no singularities of $v(x)$ at all.

In order to make this argument work, we show in the sequel that $\operatorname{grad} u_1(x)$ in fact has only isolated singularities of negative index. To this end let us first consider a nonvanishing homogeneous polynomial $p(x)$, $x \in \mathbb{R}^2$, of degree $n \geq 2$, which satisfies $\Delta p = 0$ (Δ the Laplacian). Its gradient field $\operatorname{grad} p(x)$ has a singularity at zero. We claim that its index is negative. To see this, identify $\mathbb{R}^2$ with the complex plane $\mathbb{C}$ and the point $x = (x_1, x_2)$ with $z = x_1 + ix_2$. Since $\Delta p = 0$, p is the real part of some holomorphic homogeneous polynomial. Thus, $p(z) = \operatorname{Re}(az^n)$ for a suitable $a \in \mathbb{C}$ and we obtain that

$$\operatorname{grad} p(z) = na\bar{z}^{n-1}. \tag{6.4}$$

Hence zero is an isolated singularity with index $1 - n$.

We carry over this result to the field $\operatorname{grad} u_1(x)$. Assume that $\operatorname{grad} u_1(x)$ has a singularity at zero. $u_1(x) - u_1(0)$ cannot vanish at zero of infinite order, since otherwise u_1 would be constant ([3] contains the relevant references).

Thus, applying results from [3], $u(x) - u(0)$ may be approximated around zero by a nonvanishing homogeneous polynomial p of degree $n \geq 1$, which satisfies the equation

$$\sum_{i,j=1}^{2} a_{ij}(0) \frac{\partial^2 p}{\partial x_i \, \partial x_j}(x) = 0$$

with constant coefficients. In fact, it is no loss to assume $\Delta p \equiv 0$, which can be achieved by a linear change of the coordinates [by assumption $a(0)$ is strictly positive definite]. Now the theorem of Bers states (among others) that

$$|\operatorname{grad} u_1(x) - \operatorname{grad} p(x)| = O(|x|^{n-1+\varepsilon})$$

for $|x| \to 0$ and some $\varepsilon > 0$. If $n = 1$, $\operatorname{grad} p(0) \neq 0$. On the other hand, $\operatorname{grad} u_1(0) = 0$ by assumption, which is not compatible with the above estimate. Thus, necessarily $n \geq 2$. From (6.4), $|\operatorname{grad} p(x)| = \text{const}|x|^{n-1}$. Therefore, it follows from the last estimate that $\operatorname{grad} u_1(x) \neq 0$ for small $|x| \neq 0$ and

$$\frac{\operatorname{grad} u_1(x)}{|\operatorname{grad} u_1(x)|} - \frac{\operatorname{grad} p(x)}{|\operatorname{grad} p(x)|} = O(|x|^{\varepsilon}).$$

It follows that the singularity of $\operatorname{grad} u_1$ at zero is isolated and has the same index as $\operatorname{grad} p$. Of course this works for any singularity. Thus, every singularity of $\operatorname{grad} u_1$ is isolated and has negative index. Consequently, there are no singularities at all:

$$\operatorname{grad} u_1(x) \neq 0 \qquad \forall x \in \mathbb{R}^2.$$

(iii) The rest is now rather straightforward. We have just shown $\xi^* \cdot Du(x) \cdot \xi > 0$ for $\xi = (1,0)^*$. Applying a rotation to $\mathbb{R}^2$, the same obviously is true for any ξ of length 1 such that $Du(x)$ has full rank everywhere and u is a regular mapping.

It remains to show global injectivity. Consider the set K of all points z, such that $z = u(x) = u(x')$ for some $x \neq x'$. K is bounded, since $Du(x) \to \text{Id}$ at infinity. Due to regularity, K is an open set (one can find disjoint neighbourhoods of x and x', which both are mapped onto the same neighbourhood of z). Finally let $z_n = u(x_n) = u(x'_n) \in K$ be convergent, with limit z. It is no loss to assume that x_n and x'_n converge to x and x' (say). Since u is locally injective, $x \neq x'$ and $z = u(x) = u(x') \in K$. Thus K is closed too and, consequently, the empty set. □

7. Uniqueness.

The final step is to prove the following proposition:

PROPOSITION 7.1. *Let $d = 2$ and let Assumptions A and E be satisfied. Then there is (up to translations) at most one global C^2-diffeomorphism u from $\mathbb{R}^2$ onto $\mathbb{R}^2$ such that $Lu \equiv 0$ and as $|x| \to \infty$,*

$$|u(x) - x| = o(|x|),$$
$$|Du(x) - \text{Id}| = o(1).$$

For the proof we shall use similar ideas as in Section 5. However, this time it is possible to remove the drift term in our diffusion, which facilitates the situation considerably.

PROOF OF PROPOSITION 7.1. (i) Let us first assume that the drift vector $b(x)$ vanishes everywhere. Furthermore, let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be such that $Lf \equiv 0$ and

$$f(x) = o(|x|).$$

We like to show that then $f(y) - f(z)$ is zero for all $y, z \in \mathbb{R}^2$ such that f is constant. It is no loss to assume that the second components of y and z are identical (which can always be achieved by a rotation). As in Section 4, we construct a suitable coupling. To this end let B, B', B'' be three independent standard Brownian motions on $\mathbb{R}$ and define diffusions (Y_t) and (Z_t) by

$$dY = \gamma(Y)\cdot\begin{pmatrix} dB' \\ dB \end{pmatrix}, \qquad Y_0 = y,$$

$$dZ = \gamma(Z)\cdot\begin{pmatrix} dB'' \\ dB \end{pmatrix}, \qquad Z_0 = z,$$

with

$$\gamma(x) = \begin{pmatrix} \dfrac{(\det a(x))^{1/2}}{a_{22}(x)} & \dfrac{a_{12}(x)}{a_{22}(x)} \\ 0 & 1 \end{pmatrix}.$$

Then

$$\gamma(x)\cdot\gamma(x)^* = a_{22}^{-1}(x)a(x).$$

Thus Y and Z are versions of X, up to a time change. For the second components Y_2 and Z_2 of Y and Z we have $d(Y_2 - Z_2) = 0$. Since the second components of y and z are assumed to coincide, $Y_2 = Z_2$. The distance

$$D = |Y_1 - Z_1| = |Y - Z|$$

of the first component of Y and Z obeys (as long as $D > 0$) the stochastic equation

$$dD = s(Y, Z)\, dB_D$$

for a suitable standard Brownian motion B_D, with

$$s^2(y,z) = \frac{\det a(y)}{a_{22}(y)^2} + \frac{\det a(z)}{a_{22}(z)^2} + \left(\frac{a_{12}(y)}{a_{22}(y)} - \frac{a_{12}(z)}{a_{22}(z)}\right)^2.$$

In view of Assumption E, $s^2(y,z)$ is bounded away from zero on $\mathbb{R}^4$ (cf. the proof of Lemma 4.1).

Let now $c > 0$ and $r > |y|, |z|$ and consider the cone K of points (x, d) in $\mathbb{R}^2 \times \mathbb{R}$ such that

$$d \ge 0 \quad \text{and} \quad |x - y| \le r - cd.$$

Also let

$$h(Y_t, D_t) = 1 + r^{-2}\big(|Y_t - y|^2 - (r - cD_t)^2\big).$$

Then

$$dh(Y,D) = d(\text{martingale}) + r^{-2}\left[\frac{\operatorname{trace} a(Y)}{a_{22}(Y)} - c^2 s^2(Y,Z)\right] dt$$

and, in view of Assumption E, $h(Y,D)$ is a supermartingale if only c is large enough. Let further

$$\rho = \inf\{t \geq 0 \colon (Y_t, D_t) \in \partial K\}.$$

Again $h(Y_\rho, D_\rho) \geq 0$ and if $D_\rho \neq 0$, $h(Y_\rho, D_\rho) = 1$. If r is large enough, $(Y_0, D_0) \in K$ such that by means of optional stopping,

$$\begin{aligned}\Pr_{y,z}(D_\rho \neq 0) &\leq E_{y,z} h(Y_\rho, D_\rho) \leq h(y, |y-z|)\\ &\leq 2\frac{c|y-z|}{r}.\end{aligned}$$

Now, since $Lf \equiv 0$, the usual optional stopping argument yields

$$f(y) - f(z) = E_{y,z}\big(f(Y_\rho) - f(Z_\rho)\big).$$

Furthermore, $|Y_\rho| \leq |Y_\rho - y| + |y| \leq r + |y|$, $D_\rho \leq r/c$ and $|Z_\rho| \leq |Y_\rho| + D_\rho \leq r(1 + 1/c) + |y|$. Also $Y_\rho = Z_\rho$ if $D_\rho = 0$. Thus,

$$f(y) - f(z) = o(r) \cdot \Pr_{y,z}(D_\rho \neq 0) = o(1)$$

if $r \to \infty$. Thus $f(y) = f(z)$ as we have claimed.

(ii) Now we allow an arbitrary drift component. Let u and $\bar{u}$ be two C^2-diffeomorphisms with the required properties and let $w = \bar{u} \circ u^{-1}$. Differentiating $\bar{u} = w \circ u$ we obtain

$$L\bar{u}(x) = \tilde{L}w(u(x)) + Dw(u(x)) \cdot Lu(x)$$

(L and $\tilde{L}$ are applied componentwise) with

$$\tilde{L} = \frac{1}{2}\sum_{i,j} \tilde{a}_{ij}(x) \frac{\partial^2}{\partial x_i\, \partial x_j},$$

$$\tilde{a}(u(x)) = \big(\tilde{a}_{ij}(u(x))\big) = Du(x) \cdot a(x) \cdot Du(x)^*.$$

So $Lu \equiv L\bar{u} \equiv 0$ implies $\tilde{L}w \equiv 0$. Furthermore, in view of the regularity properties of u, $\tilde{a}(x)$ inherits uniform ellipticity Assumption E from $a(x)$ (and of course Assumption A). Since $\tilde{L}$ has no first order component, also $\tilde{L}(w - \mathrm{id}) \equiv 0$. Furthermore, $|w(x) - x| = o(|x|)$, which follows from the corresponding property of u and $\bar{u}$. Hence, applying part (i), we obtain $w(x) - x = w(0)$ or $\bar{u}(x) = u(x) + w(0)$. This is the desired result. □

APPENDIX

Here we show that (4.1) and (4.2) may be solved weakly, which means that for every starting position there is a solution on a suitable probability space. Equivalently we may solve the corresponding "martingale problem." [Section

V.19 in [21] contains a nice exposition; cf. also Theorem V (20.1).] Let us quickly recall this approach. Let $\Omega = \mathscr{C}(\mathbb{R}^+, \mathbb{R}^d)$, (x_t) the canonical process [i.e., $x_t(\omega) = \omega(t)$, $\omega \in \Omega$] and $(\mathscr{F}_t)$ the canonical filtration on Ω ($\mathscr{F}_t = \sigma(x_s, s \le t)$). Let $\alpha_t = (\alpha_{ij}(t))$ and $\beta_t = (\beta_i(t))$ be progressively measurable processes on Ω with values in the nonnegative definite $d \times d$ matrices, respectively, in $\mathbb{R}^d$. Let

$$Lf_t = \frac{1}{2} \sum_{i,j} \alpha_{ij}(t) \frac{\partial^2 f}{\partial x_i \, \partial x_j}(x_t) + \sum_i \beta_i(t) \frac{\partial f}{\partial x_i}(x_t)$$

for $f \in \mathscr{C}(\mathbb{R}^+, \mathbb{R}^d)$. Then a family P_x, $x \in \mathbb{R}^d$, of probability measures on $(\Omega, (\mathscr{F}_t))$ is said to solve the *martingale problem* for (α_t, β_t) if $P_x(x_0 = x) = 1$ and

$$C_t^f = f(x_t) - \int_0^t Lf_s \, ds$$

in an $(\mathscr{F}_t)$-martingale under P_x, for every x and every $\mathscr{C}^\infty$-function f with compact support. More generally, we say that the probability measures $P_{x,s}$, $x \in \mathbb{R}^d$, $s \ge 0$, solve the martingale problem for (α_t, β_t) if $P_{x,s}(x_s = x) = 1$ and $C_t^f - C_{t \wedge s}^f$ is an $(\mathscr{F}_t)$-martingale.

PROPOSITION A.1. *We assume*:

(i) *There is an open set $D \subset \mathbb{R}^d$ as well as vectors $\beta(x)$ and nonnegative definite matrices $\alpha(x)$ for each $x \in D$ which depend continuously on x and are bounded on D.*

(ii) *There are vectors $\tilde\beta(x)$ and nonnegative definite matrices $\tilde\alpha(x)$ for each $x \in \mathbb{R}^d$ such that there is a solution $P_{x,s}$ of the martingale problem belonging to $(\tilde\alpha(x_t), \tilde\beta(x_t))$, measurable with respect to (x, s).*

Let $\tau = \inf\{t\colon x_t \notin D\}$ and

$$\alpha_t = \begin{cases} \alpha(x_t), & t < \tau, \\ \tilde\alpha(x_t), & t \ge \tau, \end{cases} \qquad \beta_t = \begin{cases} \beta(x_t), & t < \tau, \\ \tilde\beta(x_t), & t \ge \tau. \end{cases}$$

Then there is a solution of the martingale problem belonging to (α_t, β_t).

To apply this result to our coupling in Section 4, we choose $d = 4$, $D = \{(y,z) \in \mathbb{R}^2 \times \mathbb{R}^2\colon y \ne z\}$,

$$\tilde\beta(y,z) = \begin{pmatrix} b(y) \\ b(z) \end{pmatrix}, \qquad \tilde\alpha(y,z) = \begin{pmatrix} a(y) & 0 \\ 0 & a(z) \end{pmatrix}$$

and $\alpha(y,z)$, $\beta(y,z)$ according to (4.1). The validity of assumption (ii) above is then settled under Assumption E by the Stroock–Varadhan theorem (cf. Theorem 7.2.1 in [23]).

PROOF OF PROPOSITION A.1. Let Lf_t and C_t^f be as above. For $\omega \in \Omega$ with $\tau(\omega) < \infty$ define $Q_\omega = P_{x_{\tau(\omega)}, \tau(\omega)}$ such that, by assumption (ii), $C_t^f - C_{t \wedge \tau(\omega)}^f$ is an $(\mathscr{F}_t)$-martingale under Q_ω. It is then enough to construct probability measures Q_x, $x \in \mathbb{R}^d$, such that $Q_x(x_t = x) = 1$ and $C^f_{t \wedge \tau}$ is a $(\mathscr{F}_t)$-martingale under Q_x. This is the content of Theorem 6.1.2 in [23]. (The proof of this theorem works without any change for a stopping time τ, which may be infinite too.)

To construct Q_x, choose for every natural number n and $x \in \mathbb{R}^d$ a vector $\beta_n(x)$ and a nonnegative definite matrix $\alpha_n(x)$, bounded and continuous in x such that

$$\alpha_n(x) = \alpha(x), \qquad \beta_n(x) = \beta(x), \quad \text{if } \operatorname{dist}(x, D^c) \geq \frac{1}{n}.$$

Then there are probability measures Q^n_x which solve the martingale problem for $(\alpha_n(x_t), \beta_n(x_t))$ [cf. [21], Theorem V (23.5)]. However, $\alpha_n(x_t) = \alpha_t$ and $\beta_n(x_t) = \beta_t$ up to the moment

$$\tau_n = \inf\left\{t : \operatorname{dist}(x_t, D^c) \leq \frac{1}{n}\right\}.$$

Therefore, $C^f_{t \wedge \tau_n}$ is an $(\mathscr{F}_t)$-martingale under Q^m_x if $m \geq n$. Another way to say this is that

$$\int \left(C^f_{t \wedge \tau_n} - C^f_{s \wedge \tau_n}\right) \psi \, dQ^m_x = 0$$

for any $0 \leq s < t$ and any bounded, continuous $\psi \colon \Omega \to \mathbb{R}$, measurable with respect to $\mathscr{F}_s$. We want to take the limit $m \to \infty$. $(Q^m_x)_m$ is in fact a tight sequence [cf. [21], Chapter V.23, in particular Lemma (23.2)] such that it contains a weakly convergent subsequence with limit Q_x (say). Now $C^f_{t \wedge \tau_n}$ is continuous and bounded in ω. Thus, letting $m \to \infty$,

$$\int \left(C^f_{t \wedge \tau_n} - C^f_{s \wedge \tau_n}\right) \psi \, dQ_x = 0.$$

Also $\tau_n \uparrow \tau$ and $C^f_{t \wedge \tau_n} \to C^f_{t \wedge \tau}$ as $n \to \infty$ and, consequently,

$$\int \left(C^f_{t \wedge \tau} - C^f_{s \wedge \tau}\right) \psi \, dQ_x = 0.$$

This means that $(C^f_{t \wedge \tau})$ is an $(\mathscr{F}_t)$-martingale under Q_x, and the proof is finished. □

Acknowledgments. Thanks to Pawel Kröger, who made me aware of the important paper of Lipman Bers, and to the anonymous referees for their very helpful comments.

REFERENCES

[1] Arnol'd, V. I. (1973). *Ordinary Differential Equations*. MIT Press.

[2] Bernard, A., Campbell, E. A. and Davie, A. M. (1979). Brownian motion and generalized analytic and inner functions. *Ann. Inst. Fourier* **29** 207–228.

[3] Bers, L. (1955). Local behaviour of solutions of general linear elliptic equations. *Comm. Pure Appl. Math.* **8** 473–496.

[4] Chen, M. F. and Li, S.-F. (1989). Coupling methods for multidimensional diffusion processes. *Ann. Probab.* **17** 151–177.

[5] CRANSTON, M. (1991). Gradient estimates on manifolds using coupling. *J. Funct. Anal.* **99** 110–124.
[6] DURRETT, R. (1984). *Brownian Motion and Martingales in Analysis*. Wadsworth, Belmont, CA.
[7] EELLS, J. and LEMAIRE, L. (1978, 1988). A report on harmonic maps. *Bull. London Math. Soc.* **10** 1–68 (1978); Another report on harmonic maps. *Bull. London Math. Soc.* **20** 385–524.
[8] FRIEDMANN, A. (1975). *Stochastic Differential Equations and Applications* **I**. Academic Press, New York.
[9] GILBARG, D. and SERRIN, J. (1956). On isolated singularities of solutions of second order elliptic differential equations. *J. Analyse Math.* **4** 309–340.
[10] GILBARG, D. and TRUDINGER, N. S. (1977). *Elliptic Partial Differential Equations of Second Order*. Springer, Berlin.
[11] GRIFFEATH, D. (1978). *Coupling Methods for Markov Processes* 1–43. Academic Press, New York.
[12] HAS'MINSKII, R. Z. (1980). *Stochastic Stability of Differential Equations*. Sijthoff and Noordhoff, Alphen.
[13] KARP, L. (1993). Generalized Newton potential and its applications. *J. Math Anal. Appl.* **174** 480–497.
[14] KENDALL, W. S. (1987). Stochastic differential geometry: an introduction. *Acta Appl. Math.* **9** 29–60.
[15] KENDALL, W. S. (1988). Martingales on manifolds and harmonic maps. In *The Geometry of Random Motion* (M. Pinsky and R. Durrett, eds.) 121–157. Amer. Math. Soc., Providence, RI.
[16] KERSTING, G. (1989). Asymptotic properties of solutions of multidimensional SDE. *Probab. Theory Related Fields* **82** 187–211.
[17] KUNITA, H. (1969). Absolute continuity of Markov processes and generators. *Nagoya Math. J.* **36** 1–26.
[18] LEGALL, J. F. and YOR, M. (1986). Étude asymptotique de certain mouvements browniens complex avec drift. *Probab. Theory Related Fields* **71** 183–229.
[19] PINSKY, R. (1991). A probabilistic approach to a theorem of Gilbarg and Serrin. *Israel J. Math.* **74** 1–12.
[20] ROGERS, L. C. G. and LINDVALL, T. (1986). Coupling of multidimensional diffusions by reflection. *Ann. Probab.* **14** 860–872.
[21] ROGERS, L. C. G. and WILLIAMS, D. (1987). *Diffusions, Markov Processes, and Martingales*. Wiley, Chichester.
[22] SHAPIRO, H. S. (1993). A weighted L^1 estimate for the Laplace operator. *Fourier Analysis and its Applications*. Kluwer, Dordrecht.
[23] STROOCK, D. W. and VARADHAN, S. R. S. (1979). *Multidimensional Diffusion Processes*. Springer, Berlin.
[24] VENTCEL, A. D. and FREIDLIN, M. I. (1972). Some problems concerning stability under random perturbations. *Theory Probab. Appl.* **17** 269–283.
[25] VERETENNIKOV, A. JU. (1981). On strong solutions and explicit formulas for solutions of stochastic integral equations. *Math. USSR-Sb.* **39** 387–403.
[26] ZVONKIN, A. K. (1974). A transformation of the phase space of a diffusion process that removes the drift. *Math. USSR-Sb.* **22** 129–149.

FACHBEREICH MATHEMATIK
J. W. GOETHE–UNIVERSITÄT
POSTFACH 11 19 32
D-60054 FRANKFURT
GERMANY
E-MAIL: kersting@mi.informatik.uni-frankfurt.de