Lecture Notes on Kinetic Models for Waves in Random Media

Guillaume Bal [1]   Tomasz Komorowski[2]   Lenya Ryzhik[3]

September 26, 2006

[1]Department of Applied Physics and Applied Mathematics, Columbia University, New York NY, 10027; gb2030@columbia.edu
[2]Institute of Mathematics, UMCS, pl. Marii Curie Skłodowskiej 1, 20-031 Lublin, Poland; e-mail: komorow@hektor.umcs.lublin.pl
[3]Department of Mathematics, University of Chicago, Chicago IL, 60637; ryzhik@math.uchicago.edu

Contents

Chapter 1

Random geometric optics I: short time asymptotics

1.1 Rays without the wave equation – the formal theory

We consider in this section the very basic ray theory in a random emdium without any references to the wave equation – this material is based on the classical paper by J.B. Keller [38].

1.1.1 Perturbative ray theory

Fermat's principle postulates that light goes from point A to B as fast as possible. Such fastest path is called a ray connecting points A and B. The medium in which light is propagating is described in terms of the local speed of light $c(x)$. Let Γ be a ray from A to B, then

$$\int_\Gamma \frac{dl}{c(X)} = \inf_\gamma \int_\gamma \frac{dl}{c(X)}.$$

Here the infimum is taken over all continuous curves γ connecting A and B. Equivalently, parameterizing γ by $x(t) = (x_1(t), x_2(t), x_3(t))$, $0 \le t \le 1$, we need to minimize the functional

$$\int_0^1 n(x(s))|\dot{x}(s)|ds \tag{1.1}$$

with $n(x) = c_0/c(x)$ being the refractive index. Here $c_0 = \text{const}$ is a reference speed that is some typical speed of propagation in the medium. This will be sometimes formalized by requiring that $n(x)$ does not deviate from $n_0 = 1$ too much but that is not required a priori. The Euler-Lagrange equations for the functional (1.1) are

$$\frac{d}{dt}\frac{\partial F}{\partial \dot{x}_j} - \frac{\partial F}{\partial x_j} = 0$$

with $F(x, \dot{x}) = n(x(s))|\dot{x}(s)|$. This may be rewritten as

$$\frac{d}{dt}\left(\frac{n\dot{x}_j}{|\dot{x}|}\right) - |\dot{x}|\frac{\partial n}{\partial x_j} = 0.$$

Let θ be the unit vector along the ray: $\theta = \dot{x}/|\dot{x}|$, then the above equations take the form

$$\frac{d}{dt}(n\theta) - |\dot{x}|\nabla n = 0. \tag{1.2}$$

A convenient way to parameterize the path is to use the arclength parameter l along the curve $x(t)$, then $dl = |\dot{x}|dt$ and we obtain the ray equation

$$\frac{d}{dl}(n\theta) - \nabla n = 0. \tag{1.3}$$

It may be also written as

$$\frac{d}{dl}\left(n\frac{dx}{dl}\right) - \nabla n = 0. \tag{1.4}$$

Equation (1.4) should be supplemented by the initial conditions:

$$x(0) = 0,\ \frac{dx}{dl}(0) = \theta_0,\ |\theta_0| = 1. \tag{1.5}$$

This is the fundamental equation of the ray optics that describes the geometry of rays connecting different points in an inhomogeneous medium. Observe that if $n(x) = \text{const}$, then $d\theta/dl = 0$ and the direction θ doesn't change along the ray. Therefore rays in a homogeneous medium are straight lines. Similarly, if the medium is layered, that is, the refraction variable depends only on the variable x_1: $n = n(x_1)$ then rays that point initially in the direction of x_1 are straight lines – this also follows immediately from (1.5) with $n = n(x_1)$ and the initial data $\theta_0 = e_1$.

Let us now consider the case when index of refraction deviates slightly from unity:

$$n(x) = 1 + \varepsilon\mu(x).$$

We assume that ε is a small parameter: $\varepsilon \ll 1$ and employ the formal perturbation theory to determine the perturbed path $x(l,\varepsilon)$ expanding it as

$$x(l,\varepsilon) = x_0(l) + \varepsilon x_1(l) + \varepsilon^2 x_2(l) + \dots$$

We insert this expansion in the ray equations (1.4) and get in the order ε^0:

$$\frac{d^2x_0}{dl^2} = 0$$

so that $x_0(l) = l\theta_0$. The first order correction in ε is determined by the equation

$$\frac{d^2x_1}{dl^2} = \nabla\mu(x_0) - \left(\frac{dx_0}{dl}\cdot\nabla\mu(x_0)\right)\frac{dx_0}{dl} \tag{1.6}$$

with the initial condition $x_1(0) = dx_1/dl(0) = 0$. The right side of (1.6) is the component of $\nabla\mu$ normal to θ_0. We will denote it by $\nabla_\perp\mu$ below. The solution of (1.6) is given by

$$x_1(l) = \int_0^l (l-s)\nabla_\perp\mu(\theta_0 s)ds. \tag{1.7}$$

It follows that $(x_1 \cdot x_0) = 0$ – this is typical for a first order correction in the perturbation series. A straightforward computation using the fact that x_1 is perpendicular to θ_0 shows that the second order term x_2 satisfies a lengthy equation

$$\frac{d^2x_2}{dl^2} = (x_1\cdot\nabla_\perp)\nabla_\perp\mu(x_0) - \frac{1}{2}\nabla_\perp\mu^2 - \left(\frac{dx_1}{dl}\cdot\nabla_\perp\mu(x_0)\right)\frac{dx_0}{dl} - \left(\frac{dx_0}{dl}\cdot\nabla\mu(x_0)\right)\frac{dx_1}{dl} \tag{1.8}$$

with the initial data $x_2(0) = dx_2/dl(0) = 0$. Its solution is given by

$$x_2(l) = \int_0^l ds(l-s)\left[(x_1(s)\cdot\nabla_\perp)\nabla_\perp\mu(s\theta_0) - \frac{1}{2}\nabla_\perp\mu^2(s\theta_0) - \theta_0\int_0^s(\nabla_\perp\mu(s\theta_0)\cdot\nabla_\perp\mu(\tau\theta_0))d\tau\right.$$
$$\left. - (\theta_0\cdot\nabla\mu(s\theta_0))\int_0^s\nabla_\perp\mu(\tau\theta_0)d\tau\right]. \tag{1.9}$$

1.1.2 Weakly perturbed rays in a random medium

Expressions for the corrections x_1 and x_2 obtained above are valid for any perturbation $\mu(x)$. Let us now specify that $\mu(x)$ is a random function that has mean zero and its statistics is spatially homogeneous and isotropic:

$$\langle\mu(x)\rangle = 0, \quad \langle\mu(x)\mu(y)\rangle = R(|x-y|), \quad \langle\mu(p)\mu(q)\rangle = (2\pi)^n\hat{R}(p)\delta(p+q). \tag{1.10}$$

The correlation function $R(|x|)$ is smooth, has maximum at zero, is a decreasing and rapidly decaying function of $|x|$, and the power spectrum $\hat{R}(p)$ is its Fourier transform.

The mean ray position

Let us first compute the average ray position using expressions obtained in Section 1.1. The first order correction has mean zero: $\langle x_1\rangle = 0$ so that

$$\langle x(l)\rangle = l\theta_0 + \varepsilon^2\langle x_2\rangle + O(\varepsilon^3).$$

The expected value of of x_2 may be computed explicitly using expression (1.9). Observe that $\langle\nabla_\perp\mu^2\rangle = 0$ since $\langle\mu^2\rangle$ is constant because of spatial homogeneity. To compute the mean of the first term in (1.9) we use (1.7) and the Fourier transform, as well as the last identity in (1.10). This term becomes:

$$\left\langle\int_0^l ds\int_0^s ds_1(l-s)(s-s_1)(\nabla_\perp\mu(s_1\theta_0)\cdot\nabla_\perp)\nabla_\perp\mu(s\theta_0)\right\rangle = \int_0^l ds\int_0^s ds_1(l-s)(s-s_1)$$
$$\times\int e^{is_1p\cdot\theta_0+isq\cdot\theta_0}\left((ip-i(\theta_0\cdot p)\,\theta_0)\cdot(iq-i(\theta_0\cdot q)\theta_0)\right)[iq-i(\theta_0\cdot q)\theta_0]\,\langle\hat{\mu}(p)\hat{\mu}(q)\rangle\frac{dpdq}{(2\pi)^{2n}}$$
$$= -i\int_0^l ds\int_0^s ds_1(l-s)(s-s_1)\int e^{is_1p\cdot\theta_0-isp\cdot\theta_0}\,|p-(\theta_0\cdot p)\theta_0|^2\,(p-(\theta_0\cdot p)\theta_0)\hat{R}(p)\frac{dp}{(2\pi)^n}$$

We claim that the integral in p vanishes. To see that we choose a coordinate system such that $\theta_0 = e_n$ is the unit vector in the direction of x_n. The integral in p becomes

$$\int e^{i(s_1-s)p_n}|p'|^2p'\hat{R}(p)\frac{dp}{(2\pi)^n} = 0, \quad p' = (p_1, p_2, \ldots, p_{n-1}, 0),$$

because $\hat{R}$ depends only on $|p|$ so that $\hat{R}(p', p_n) = \hat{R}(-p', p_n)$. Similarly one may show that the average of the last term in (1.9) vanishes. Indeed, it is equal to

$$
\begin{aligned}
&-\left\langle \int_0^l (l-s)(\theta_0 \cdot \nabla\mu(s\theta_0)) \int_0^s \nabla_\perp \mu(\tau\theta_0) d\tau ds \right\rangle \\
&= -\int_0^l \int_0^s (l-s) \int e^{isp\cdot\theta_0 + i\tau q\cdot\theta_0} \langle \hat{\mu}(p)\hat{\mu}(q)\rangle (\theta_0 \cdot (ip)) \left[iq - i(q\cdot\theta_0)\theta_0\right] \frac{dpdq}{(2\pi)^n} d\tau ds \\
&= -\int_0^l \int_0^s (l-s) \int e^{isp\cdot\theta_0 - i\tau p\cdot\theta_0} \hat{R}(p)(\theta_0 \cdot p) \left[p - (p\cdot\theta_0)\theta_0\right] \frac{dp}{(2\pi)^n} d\tau ds.
\end{aligned}
$$

With the same choice of coordinates such that $\theta_0 = e_n$ is along the p_n-variable and $p' = (p_1, \ldots, p_{n-1}, 0)$ the integral in p above is

$$
\int e^{i(s-\tau)p_n} \hat{R}(p) p_n p' \frac{dp}{(2\pi)^n} = 0
$$

as $\hat{R}(p', p_n) = \hat{R}(-p', p_n)$ because $R(x)$ is isotropic.

Thus only the third term has a non-zero mean and we obtain

$$
\langle x(l)\rangle = l\theta_0 - \varepsilon^2\theta_0 \int_0^l ds(l-s) \int_0^s d\tau \left\langle (\nabla_\perp\mu(s\theta_0) \cdot \nabla_\perp\mu(\tau\theta_0))\right\rangle + O(\varepsilon^3).
$$

The second term may be evaluated using the Fourier transform as above:

$$
\langle \nabla_\perp\mu(s\theta_0) \cdot \nabla_\perp\mu(\tau\theta_0)\rangle = \int e^{i(s-\tau)p\cdot\theta_0} (p - (\theta_0\cdot p)\theta_0)^2 \hat{R}(|p|) \frac{dp}{(2\pi)^n} = -\Delta_\perp R((s-\tau)\theta_0).
$$

Here $\Delta_\perp$ is the Laplacian in the direction transverse to θ_0: with our usual choice of θ_0 along the coordinate x_n we have

$$
\Delta_\perp R = \sum_{j=1}^{n-1} \frac{\partial^2 R}{\partial x_j^2}.
$$

When $R(|x|)$ is radially symmetric this is

$$
\Delta_\perp R(x) = \sum_{j=1}^{n-1} \frac{\partial}{\partial x_j}\left[\frac{R'(|x|)x_j}{|x|}\right] = \sum_{j=1}^{n-1} \left(\frac{R''(|x|)x_j^2}{|x|^2} + \frac{R'(|x|)}{|x|} - \frac{R'(|x|)x_j^2}{|x|^3}\right),
$$

and in particular

$$
\Delta_\perp R(r\theta_0) = \frac{(n-1)R'(r)}{r}.
$$

This leads to

$$
\langle x(l)\rangle = \theta_0 \left(l + \frac{(n-1)\varepsilon^2}{2} \int_0^l (l-r)^2 \frac{R'(r)}{r} \right) dr + O(\varepsilon^3). \tag{1.11}
$$

We see that the mean location of the endpoint of a ray of length l which starts from the origin is in the direction θ_0. However, its distance from the origin is less than when the ray is not perturbed since $R'(r) < 0$. Physically this is expected because the presence of the inhomogeneities slows down the propagation as light no longer propagates along a straight line.

The mean square fluctuations

Let us denote by ρ deviation of the ray from the straight line $x = \theta_0 l$: $\rho(l) = x_\varepsilon(l) - \theta_0 l$. Then we have

$$\langle \rho(l)^2 \rangle = \varepsilon^2 \langle x_1^2 \rangle + o(\varepsilon^2).$$

We set now space dimension $n = 3$ and compute the above average in the same manner as before:

$$\langle x_1^2(l) \rangle = -\int_0^l \int_0^l (l-s)(l-\tau)\Delta_\perp R(|s-\tau|) d\tau ds = -\int_0^l R'(r) \left[\frac{2r^2}{3} - 2l^2 + \frac{4l^3}{3r}\right] dr.$$

For l large compared to the correlation length a the last term above dominates so that

$$\langle \rho(l)^2 \rangle \approx -\frac{4\varepsilon^2 l^3}{3} \int_0^\infty \frac{R'(r)}{r} dr + O(\varepsilon^3) \text{ for } l \gg a. \tag{1.12}$$

Similarly one may compute the average deviation α of the direction of the ray from the mean direction θ_0:

$$\langle \alpha^2(l) \rangle = \left\langle \left(\frac{dx}{dl} - \theta_0\right)^2 \right\rangle = \varepsilon^2 \left\langle \left|\frac{dx_1}{dl}\right|^2 \right\rangle + O(\varepsilon^3) = \varepsilon^2 \left\langle \int_0^l \int_0^l \nabla_\perp \mu(\theta_0 s) \nabla_\perp \mu(\theta_0 \tau) \right\rangle d\tau ds$$

$$= -2\varepsilon^2 \int_0^l \int_0^l \frac{R'(|\tau - s|)}{|\tau - s|} d\tau ds = 4\varepsilon^2 \left[R(l) - R(0) - l \int_0^l \frac{R'(r)}{r} dr\right] + O(\varepsilon^3).$$

For l large compared to the correlation length a this becomes

$$\langle \alpha^2(l) \rangle \approx -4\varepsilon^2 l \int_0^\infty \frac{R'(r)}{r} dr + O(\varepsilon^3) \text{ for } l \gg a. \tag{1.13}$$

Expressions (1.12) and (1.13) may be written as

$$\langle \rho^2(l) \rangle \approx \frac{1}{3} D l^3 + O(\varepsilon^3) \tag{1.14}$$

and

$$\langle \alpha^2(l) \rangle \approx D l + O(\varepsilon^3). \tag{1.15}$$

Here we introduced the ray diffusion coefficient

$$D = -4\varepsilon^2 \int_0^\infty \frac{R'(r)}{r} dr. \tag{1.16}$$

Expressions (1.14) and (1.15) may also be obtained by treating the ray direction $\alpha(l)$ as a Brownian motion with the diffusion coefficient D and $\rho(l)$ as its time integral:

$$d\alpha = \sqrt{D} dB, \quad d\rho = \alpha(l) dl.$$

Here $B(l)$ is the standard Brownian motion. Then a simple calculation shows that

$$\langle \rho(l)^2 \rangle = D \left\langle \int_0^l \int_0^l B(s) B(s') ds' ds \right\rangle = 2D \left\langle \int_0^l \int_0^{s'} B(s) B(s') ds ds' \right\rangle = 2D \int_0^l \int_0^s s ds ds' = \frac{Dl^3}{3}.$$

However, the a priori assumption that ray direction may be described in terms of such Markov process is not easy to justify unlike the derivation presented above. Nevertheless, this concept is important and the ray direction does behave as a Markov process in a certain asymptotic limit that we will discuss in the rest of this chapter.

1.1.3 Random Liouville equations: small time formal asymptotics

Reduction to a time-dependent stochastic accelaration problem

In order to make the above discussion of the diffusive ray behavior somewhat more careful (albeit not yet rigorous) we consider the Liouville equations in phase space

$$\frac{\partial \phi}{\partial t} + c(x)\hat{k} \cdot \nabla_x \phi - |k|\nabla c(x) \cdot \nabla_k \phi = 0 \tag{1.17}$$

with the speed $c(x) = 1 + \delta\mu(x)$. Here $\mu(x)$ is a spatially homogeneous random process with the correlation function as in (1.10). Then (1.17) becomes

$$\frac{\partial \phi}{\partial t} + [1 + \delta\mu(x)]\hat{k} \cdot \nabla_x \phi - \delta\nabla\mu(x)|k| \cdot \nabla_k \phi = 0 \tag{1.18}$$

and solutions are close to those of

$$\frac{\partial \phi}{\partial t} + \hat{k} \cdot \nabla_x \phi = 0 \tag{1.19}$$

for times $t = O(1)$. In order to see some more interesting phenomena, in particular, the ray diffusion mentioned above, we look at the bicharacteristics of (1.18):

$$\dot{X}(t) = -(1 + \delta\mu(X))\hat{K}(t), \quad \dot{K}(t) = \delta\nabla\mu(X(t))|K(t)|, \quad X(0) = x_0, \quad K(0) = k_0. \tag{1.20}$$

It is convenient to re-write this system in terms of the unit vector $\hat{K}(t) = K(t)/|K(t)|$ using the relation

$$\frac{d\hat{K}}{dt} = \frac{\dot{K}}{|K|} - \frac{(K \cdot \dot{K})}{|K|^2}\hat{K} = \delta\nabla\mu(X) - \delta(\hat{K} \cdot \nabla\mu(X))\hat{K}.$$

This allows us to re-write (1.20) as

$$\frac{dX(t)}{dt} = -(1 + \delta\mu(X))\hat{K}(t), \quad \frac{d\hat{K}(t)}{dt} = \delta[\nabla\mu(X(t)) - (\hat{K} \cdot \nabla\mu(X))\hat{K}]. \tag{1.21}$$

Let us introduce the rescaled quantities $Y(t) = X(t) + \hat{k}_0 t - x_0$ and $P = (\hat{K}(t) - \hat{k}_0)/\delta^\alpha$ with $\alpha > 0$ to be chosen. Naively, one would expect that over a time T the direction $\hat{K}$ deviates from its initial value by δT which means that the trajectory deviates from $X_0(t) = -\hat{k}_0 t$ by $T \cdot \delta T = \delta T^2$. Hence we would expect that $Y(t)$ behaves non-trivially on the time scale $O(\delta^{-1/2})$. We will see, however, that because the random perturbation has mean zero, the effect takes place on a longer time scale.

In the new variables the system (1.21) becomes

$$\begin{aligned} \dot{Y}(t) &= -\delta^\alpha P(t) - \delta\mu(X(t))\hat{k}_0 - \delta^{1+\alpha}\mu(X(t))P(t), \\ \dot{P}(t) &= \delta^{1-\alpha}[\nabla\mu(X(t)) - (\hat{K} \cdot \nabla\mu(X))\hat{K}]. \end{aligned} \tag{1.22}$$

In the slow time variable $t' = \delta^\alpha t$ this is

$$\begin{aligned} \frac{d\tilde{Y}(t')}{dt'} &= -\tilde{P}(t') - \delta^{1-\alpha}\mu\left(x_0 - \frac{k_0 t'}{\delta^\alpha} + \tilde{Y}(t')\right)\hat{k}_0 - \delta\mu\left(x_0 - \frac{k_0 t'}{\delta^\alpha} + \tilde{Y}(t')\right)\tilde{P}(t'), \\ \frac{d\tilde{P}(t')}{dt'} &= \delta^{1-2\alpha}\left[I - (\hat{k}_0 + \delta^\alpha\tilde{P}(t')) \otimes (\hat{k}_0 + \delta^\alpha\tilde{P}(t'))\right]\nabla\mu\left(x_0 - \frac{k_0 t'}{\delta^\alpha} + \tilde{Y}(t')\right) \end{aligned} \tag{1.23}$$

with $\tilde{Y}(t') = Y(t'/\delta^\alpha)$ and $\tilde{P}(t') = P(t'/\delta^\alpha)$. We choose $\alpha = 2/3$ so that $\delta^{1-2\alpha} = \delta^{-\alpha/2}$ and introduce $\varepsilon = \delta^{1/3}$:

$$\begin{aligned}\frac{d\tilde{Y}(t')}{dt'} &= -\tilde{P}(t') - \varepsilon\mu\left(x_0 - \frac{k_0 t'}{\varepsilon^2} + \tilde{Y}(t')\right)\hat{k}_0 - \varepsilon^3\mu\left(x_0 - \frac{k_0 t'}{\varepsilon^2} + \tilde{Y}(t')\right)\tilde{P}(t'), \qquad (1.24)\\ \frac{d\tilde{P}(t')}{dt'} &= \frac{1}{\varepsilon}\left[I - (\hat{k}_0 + \varepsilon^2\tilde{P}(t'))\otimes(\hat{k}_0 + \varepsilon^2\tilde{P}(t'))\right]\nabla\mu\left(x_0 - \frac{k_0 t'}{\varepsilon^2} + \tilde{Y}(t')\right).\end{aligned}$$

Let us keep only the leading order terms in (1.24). The analysis that we perform on the simplified system may be applied to the full problem as well albeit at the price of somewhat lengthier calculations that we are not willing to pay at the moment. Then (1.24) becomes (we now drop both the primes and tildes)

$$\begin{aligned}&\dot{Y}(t) = -P(t), \quad Y(0) = 0 \qquad (1.25)\\ &\dot{P}(t) = \frac{1}{\varepsilon}\left[I - (\hat{k}_0\otimes\hat{k}_0)\right]\nabla\mu\left(x_0 - \frac{k_0 t}{\varepsilon^2} + Y(t)\right), \quad P(0) = 0,\end{aligned}$$

which is the system we will study. The vector $\dot{P}(t)$ is orthogonal to k_0 for all $t \geq 0$ – hence so is $P(t)$ and thus $(Y(t)\cdot k_0) = 0$ for all $t \geq 0$ as well. This is a familiar phenomenon for the perturbation theory – the first order correction is orthogonal to the mean dispalcement. It is convenient to set $x_0 = 0$ and choose the coordinate axes so that $k_0 = e_n$, the unit vector in the direction of x_n. Then $Y(t) = (Y_1, \ldots, Y_{n-1}, 0)$, $P(t) = (P_1, \ldots, P_{n-1}, 0)$ and (1.25) may be re-written as the following system for $Z(t) = (Y_1(t), \ldots, Y_{n-1}(t))$ and $Q(t) = (P_1(t), \ldots, P_{n-1}(t))$:

$$\begin{aligned}&\dot{Z}(t) = -Q(t), \quad Z(0) = 0 \qquad (1.26)\\ &\dot{Q}(t) = \frac{1}{\varepsilon}G\left(\frac{t}{\varepsilon^2}, Z(t)\right), \quad Q(0) = 0,\end{aligned}$$

where

$$G(x_n, x') = \left(\frac{\partial\mu(x', -x_n)}{\partial x_1}, \ldots, \frac{\partial\mu(x', -x_n)}{\partial x_{n-1}}\right), \quad x' \in \mathbb{R}^{n-1}.$$

This system also happens to describe the motion of a classical particle moving in a random time-dependent force field $\varepsilon^{-1}G(t/\varepsilon^2, x)$ and is called the stochastic acceleration problem in this context. Equivalently, (1.26) describes the trajectories for the Liouville equation

$$\phi_t + k\cdot\nabla_x\phi - \frac{1}{\varepsilon}G(\frac{t}{\varepsilon^2}, x)\cdot\nabla_k\phi = 0$$

which appears in the semi-classical limit $\delta \to 0$ for the Schrödinger equation

$$i\delta\frac{\partial\psi}{\partial t} + \frac{\delta^2}{2}\Delta\psi - \frac{1}{\varepsilon}G(\frac{t}{\varepsilon^2}, x)\psi = 0, \quad \psi(0, x) = \psi_\delta^0(x),$$

with the parameter $\varepsilon > 0$ fixed.

A very formal derivation of the diffusive limit

We now describe a very formal but quick and effective way to obtain the limit of (1.26) as $\varepsilon \to 0$. Let us write the corresponding Liouville equation

$$\frac{\partial\phi}{\partial t} + q\cdot\nabla_z\phi - \frac{1}{\varepsilon}G\left(\frac{t}{\varepsilon^2}, z\right)\cdot\nabla_q\phi = 0. \qquad (1.27)$$

Instead of assuming that the random function $G(s,x)$ is as in (1.26) we make a more general hypothesis that for each $x \in \mathbb{R}^n$ the process $G(s,x) \in \mathbb{R}^n$ is stationary in time with the two-point correlation tensor

$$R_{ml}(s,x) = \langle G_m(t,x)G_p(t+s,x)\rangle.$$

We seek the solution as a multiple time scale expansion

$$\phi = \phi_0(t,z) + \varepsilon\phi_1(t,\tau,z) + \varepsilon^2\phi_2(t,\tau,z) + \dots,\ \ \tau = t/\varepsilon^2. \tag{1.28}$$

As usual in such expansions in random media we assume that the leading order term is independent of the fast variable and is deterministic. The higher order corrections are assumed to be stationary in the fast variable τ. These assumptions are typically very hard to justify rigorously – nevertheless they often provide the correct answer. We insert the expansion into (1.27) and obtain in the leading order $O(1/\varepsilon)$

$$\frac{\partial\phi_1}{\partial\tau} = G(\tau,z)\cdot\nabla_q\phi_0(t,z)$$

so that

$$\phi_1(t,z,\tau) = \chi(\tau,z)\cdot\nabla_q\phi_0(t,z) \tag{1.29}$$

with the corrector $\chi(\tau,z)$ that solves $\dot{\chi} = G(\tau,z)$. It is very convenient to introduce a regularization parameter θ that we will send to zero later and write

$$\chi_m(\tau,z) = \int_{-\infty}^{\tau} e^{\theta s}G_m(s,z)ds. \tag{1.30}$$

The terms of the order $O(1)$ in (1.27) are

$$\frac{\partial\phi_0}{\partial t} + q\cdot\nabla_z\phi_0 - G\left(\tau,z\right)\cdot\nabla_q\phi_1 + \frac{\partial\phi_2}{\partial\tau} = 0.$$

We take the expectation of this equation using the fact that ϕ_0 is deterministic and argue that because ϕ_2 is stationary in τ we have

$$\left\langle\frac{\partial\phi_2}{\partial\tau}\right\rangle = 0.$$

With these two closure assumptions we obtain

$$\frac{\partial\phi_0}{\partial t} + q\cdot\nabla_z\phi_0 = \langle G\left(\tau,z\right)\cdot\nabla_q\phi_1\rangle.$$

The term on the right side is computed explicitly using expression (1.29)-(1.30) for ϕ_1:

$$\langle G\left(\tau,z\right)\cdot\nabla_q\phi_1\rangle = \left\langle G_m\left(\tau,z\right)\frac{\partial}{\partial q_m}\left[\int_{-\infty}^{\tau} e^{\theta s}G_p(s,z)ds\frac{\partial\phi_0}{\partial q_p}\right]\right\rangle$$
$$= \int_{-\infty}^{\tau} e^{\theta s}R_{mp}(s-\tau,z)ds\frac{\partial^2\phi_0}{\partial q_m\partial q_p} \to \int_{-\infty}^{0} R_{mp}(s,z)ds\frac{\partial^2\phi_0}{\partial q_m\partial q_p} \text{ as } \theta\to 0.$$

Therefore, the function $\phi_0(t,q,z)$ satisfies a degenerate parabolic equation

$$\frac{\partial\phi_0}{\partial t} + q\cdot\nabla_z\phi_0 = D_{mp}(z)\frac{\partial^2\phi_0}{\partial q_m\partial q_p} \tag{1.31}$$

with the symmetrized diffusion coefficient

$$D_{mp}(z) = \frac{1}{2}\left[\int_{-\infty}^{0} R_{mp}(s,z)ds + \int_{-\infty}^{0} R_{pm}(s,z)ds\right] = \frac{1}{2}\int_{-\infty}^{\infty} R_{mp}(s,z)ds.$$

If the statistics of $G(s,x)$ is identical for all $x \in \mathbb{R}^n$ then the diffusion matrix is constant in space. This means that in the limit $\varepsilon \to 0$ the process $Q(t)$ becomes a diffusion while $Z(t)$ is its integral in time.

Going back to the short time asymptotics for the geomtric optics we see that the rescaled deviation of the wave vector from its original value k_0 converges to a diffusion process $Q(t)$ and the deviation of the spatial position from its average $k_0 t$ converges to the time integral of $Q(t)$ on a time scale of the order $O(\delta^{-2/3})$. This time is much longer than the naive prediction $O(\delta^{-1/2})$ discussed below (1.21). Here δ is the relative size of the variations of the refraction index. This provides a formalization of the ray diffusion we have discussed in Section 1.1.2, at least for short times. It turns out that the randomization of the wave vector on the time scale $O(\delta^{-2/3})$ is related to the appearance of a caustic. kIt has been shown by B. White [61] appears on this time scale with probability one. This means that the ray approach in a random medium works only on a very short time scale as caustics appear very qucikly. On the other hand, one may follow the solutions of the Liouville equations for arbitrarily long times.

1.2 Basic facts on weak convergence in C and D

Weak convergence

We recall in this section basic facts from [11] on weak convergence of probability measures. All the proofs of the results of this section can be found there as well as a wealth of other information. Recall that a sequence of Borel measures P_n defined on a space Ω converges weakly to a Borel measure P on Ω if for every bounded continuous real function f we have

$$\int_{\Omega} f dP_n \to \int_{\Omega} f dP.$$

This condition is equivalent to the following: for every set A with $P(\partial A) = 0$ we have $P_n(A) \to P(A)$. A family F of (Borel) probability measures on Ω is relatively weakly compact if every sequence P_n of elements in F contains a weakly convergent subsequence P_{n_k} which converges weakly to a probability measure Q.

Weak convergence in C

One of the main examples we are going to discuss is the weak convergence in the space $C = C([0,T];\mathbb{R}^n)$ of continuous functions (paths) on an interval $[0,T]$. An effective way to verify weak compactness in C is provided by Prokhorov's theorem. Recall that a family F of probability measures is tight if for every $\varepsilon > 0$ there exists a compact set K so that $P(K) > 1 - \varepsilon$ for all measures $P \in F$.

Theorem 1.2.1 *If a family F is tight then it is relatively compact.*

As a corollary we have the following basic criterion for weak convergence.

Corollary 1.2.2 *Let P_n and P be probability measures on C. If the finite-dimensional distirbutions of P_n converge weakly to those of P and $\{P_n\}$ is a tight family then P_n converge weakly to P.*

It is important to note that convergence of finite-dimensional distributions in C in itself does not imply weak convergence and tightness assumption in Corollary 1.2.2 can not be dropped. Indeed, consider a sequence of piece-wise linear functions z_n which increase from 0 to 1 on the interval $[0, 1/n]$, decrease from 1 to 0 on the interval $[1/n, 2/n]$ and are equal to zero for $t \geq 2/n$. Set the measure $P_n = \delta_{z_n}$ and let $P = \delta_0$, the delta-function concentrated on the function $z = 0$. Suppose that A is a finite-dimensional subset of C, that is, there exists a finite set of times $t_1, \ldots, t_k$ so that if a path $x(t)$ lies in A then so do all paths $y(t)$ such that $x(t_i) = y(t_i)$ for all $1 \leq i \leq k$. If A is a finite-dimensional set then as soon as n is so large that $1/n < t_i$ for all $i = 1, \ldots, k$ such that $t_i > 0$ (this qualifier is needed since it is possible that some $t_i = 0$) we have $P_n A = PA$ simply because $z_n(t_j) = z(t_j)$ for all $j = 1, \ldots, k$ (including the time $t_i = 0$ if there is such an i) and thus z_n lies in A if and only if $z \in A$. On the other hand if we define $f(x) = \min[2, \|x\|]$ with the uniform norm

$$\|x\| = \sup_{0 \leq t \leq 1} |x(t)|$$

then f is by definition a bounded continuous function on C but

$$\int f dP_n = 1$$

while

$$\int f dP = 0.$$

Therefore P_n does not converge weakly to P. This example shows that convergence of finite-dimensional distributions is not sufficient for weak convergence.

The advanatage of tightness is that it is a verifiable notion by means of various moduli of continuity. The usual modulus of coninuity of a function $x(t)$, $t \in [0, 1]$ is defined as

$$w_x(\varepsilon) = \sup_{|t-s| \leq \varepsilon} |x(s) - x(t)|, \quad 0 < \varepsilon \leq 1.$$

The Arzela-Ascoli theorem implies that a set A is relatively compact in C if and only if both $\sup_{x \in A} |x(0)| < +\infty$ and

$$\lim_{\varepsilon \to 0} \sup_{x \in A} w_x(\varepsilon) = 0.$$

The following theorem (Theorem 7.3 in [11]) is the most basic criterion for tightness in C.

Theorem 1.2.3 *A sequence of probability measures P_n on C is tight if and only if the following two conditions hold: (i) for each $\eta > 0$ there exist n_0 and $a > 0$ so that*

$$P_n[x : \ x(0) \geq a] \leq \eta \text{ for all } n \geq n_0, \tag{1.32}$$

and (ii) for each $\delta > 0$ and $\eta > 0$ there exists $0 < \varepsilon < 1$ and n_0 so that

$$P_n \left[x : \ w_x(\varepsilon) \geq \delta\right] \leq \eta \text{ for all } n \geq n_0. \tag{1.33}$$

Condition (1.32) is usually easy to verify, especially so when we the measures P_n are generated by solutions of differential equations (with coefficients that depend on the parameter n) with a prescribed initial point – then $x(0)$ does not depend on n. On the other hand, verifying (1.33) is the heart of the proof of many limit theorems. Some criteria for (1.33) to hold will be given in the next section.

The space D

The cadlag functions. It is quite common that one has to deal with convergence of processes that have jumps but are "nice" otherwise. The appropriate space to work with consists of functions that have limits on the left and are continuous on the right:

$$\begin{aligned}&\text{(i) For } 0 \le t < 1 \text{ the right limit } x(t^+) = \lim_{s\to t^+} x(s) \text{ exists and } x(t) = x(t^+).\\ &\text{(ii) For } 0 < t \le 1 \text{ the left limit } x(t^-) = \lim_{s\to t^-} x(s).\end{aligned} \tag{1.34}$$

Such functions are often called cadlag functions ("continu á droite, limites á gauche").

Cadlag functions can not be too bad: for instance, it is easy to check that for any cadlag function $x(t)$ and any $\varepsilon > 0$ one can find a finite partition $0 = t_0 < t_1 < \cdots < t_n = 1$ of the interval $[0, 1]$ such that the oscillation $w_x[t_{i-1}, t_i) < \varepsilon$. Here the oscillation of the function $x(t)$ on a set S is defined as

$$w_x(S) = \sup_{s,t\in S} |x(s) - x(t)|. \tag{1.35}$$

It follows that any cadlag function $x(t)$ is uniformly bounded and, moreover, has at most countably many discontinuities since the number of points where the jump magnitude exceeds $1/n$ is finite for all $n \in \mathbb{N}$. We will continue to denote the usual uniform norm by

$$\|x\| = \sup_{0\le t\le 1} |x(t)|.$$

The Skorohod topology. The usual uniform topology is too rigid to work in the space D. If we think of functions in D as, for instance, realizations of a random jump process then we would like to think of two realizations as close even if the jumps occur not at exactly the same time but rather at close times. The uniform norm does not capture this idea. Instead, for two functions x and y in D we define the distance $d(x, y)$ as the smallest number $\varepsilon > 0$ so that we may find an increasing continuous function ("time change") $\lambda(t)$ such that $\lambda(0) = 0$, $\lambda(1) = 1$ and both

$$\sup_{t\in[0,1]} |\lambda(t) - t| < \varepsilon$$

and

$$\sup_{t\in[0,1]} |x(t) - y(\lambda(t))| = \sup_{t\in[0,1]} |x(\lambda^{-1}(t)) - y(t)| < \varepsilon. \tag{1.36}$$

This metric defines the Skorohod topology.

Let Λ be the set of increasing continuous functions $\lambda(t)$ such that $\lambda(0) = 0$, $\lambda(1) = 1$. Then a sequence $x_n(t)$ converges to $x(t)$ in the Skorohod topology in D if there exists a sequence $\lambda_n \in \Lambda$ such that $\tilde{x}_n(t) = x_n(\lambda_n(t))$ converges to $x(t)$ and $\lambda_n(t)$ converges to t – both in the uniform topology of $[0, 1]$. In particular, the usual uniform convergence implies convergence in the Skorohod topology – simply take $\lambda_n(t) = t$. Moreover, as

$$|x_n(t) - x(t)| \le |x_n(t) - x(\lambda_n(t))| + |x(\lambda_n(t)) - x(t)|, \tag{1.37}$$

it follows that $x_n(t)$ converges pointwise to $x(t)$ at the points where $x(t)$ is continuous. Since $x(t)$ is continuous for all but countably many points, Skorohod convergence implies pointwise convergence except on a countable set of points. In addition (1.37) implies that if the limit $x(t)$ is continuous on $[0, 1]$ (and hence uniformly continuous) then the Skorohod convergence implies the uniform convergence.

The problem with the above metric is that the space D is not complete under the metric d as can be seen on the following example. Let $x_n(t) = 1$ for $0 \le t \le 1/2^n$ and $x_n(t) = 0$ otherwise. Let $\lambda_n \in \Lambda$ be a (piecewise) linear function:

$$\lambda_n(t) = \frac{t}{2}$$

on the interval $[0, 1/2^n]$ and

$$\lambda_n(t) = \frac{1}{2^{n+1}} + \frac{1 - \frac{1}{2^{n+1}}}{1 - \frac{1}{2^n}} \left(t - \frac{1}{2^n} \right)$$

on the interval $[1/2^n, 1]$ so that λ_n maps $[0, 1/2^n]$ onto $[0, 1/2^{n+1}]$. Then $x_{n+1}(\lambda_n(t)) = x_n(t)$ and $|\lambda_n(t) - t| \le 1/2^{n+1}$. This means that $d(x_n, x_{n+1}) \le 1/2^{n+1}$ and therefore the sequence $x_n(t)$ is Cauchy in the metric d. On the other hand, $x_n(t)$ converges pointwise to $x(t) = 0$ for all $t > 0$. Therefore, if x_n converges in the Skorohod topology the only possible limit function is $x(t) = 0$ (because Skorohod convergence implies pointwise convergence except on a countable set). However, the distance from each $x_n(t)$ to $x = 0$ is equal to one (simply because $x(\lambda(t)) \equiv 0$ for all $\lambda \in \Lambda$ and $x_n(0) = 1$ for all n) and thus $x_n(t)$ does not converge in the Skorohod topology.

The way to make the space D complete is to introduce a different metric d_0 defined as follows. For $\lambda \in \Lambda$ define

$$\|\lambda\|_0 = \sup_{s<t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right|.$$

This means that the slopes of λ are bounded away from zero and infinity if $\|\lambda\|_0 < \infty$. The distance $d_0(x, y)$ for $x, y \in D$ is the smallest number $\varepsilon \ge 0$ so that there exists $\lambda \in \Lambda$ such that $\|\lambda\|_0 < \varepsilon$ and (1.36) holds. This is more restrictive than d: it requires that not only λ is close to identity in the uniform norm but the slopes of λ are all close to one. In particular, the above example of a non-converging Cauchy sequence involves λ_n which are not close to identity in this norm. We have the following proposition.

Proposition 1.2.4 *The metrics d and d_0 are equivalent on D in the sense that $d(x_n, x) \to 0$ if and only if $d_0(x_n, x) \to 0$. Moreover, the space D is separable under both d and d_0 and complete under d_0.*

Note that there is no contradiction in this proposition to the above example of a sequence x_n which is d-Cauchy in D but does not converge. This sequence is simply not d_0-Cauchy: $d_0(x_n, x_{n+1}) = \|\lambda_n\|_0 = \log 2$.

Compactness in D

Compactness in terms of $w'_x(\delta)$. Modulus of coninuity is not a right notion for a function in D as $w_x(\delta)$ does not vanish in the limit $\delta \to 0$. An alternative modulus which allows for jumps is defined as follows. We have mentioned that for any function $x(t) \in D$ and any $\varepsilon > 0$ one can find a finite partition $0 = t_0 < t_1 < \cdots < t_n = 1$ such that on each sub-interval the oscillation $w_x[t_{i-1}, t_i) < \varepsilon$. We say that a partition $\{t_i\}$ is δ-sparse if $t_i - t_{i-1} > \delta$ for all i. Define the modulus

$$w'_x(\delta) = \inf_{\{t_i\}} \max_{1 \le i \le n} w_x[t_{i-1}, t_i)$$

with the infimum taken over all δ-sparce partitions $\{t_i\}$. The previous argument shows that $\lim_{\delta \to 0} w'_x(\delta) = 0$ for any cadlag function $x \in D$. It is straightforward to check that we always have $w'_x(\delta) \le w_x(2\delta)$. There can be no inequality in the opposite direction because the usual modulus of conitnuity $w_x(\delta)$ does not go to zero as $\delta \to 0$ for a discontinuous function from D. However, for a continuous function $x(t)$ we do have an inequality $w_x(\delta) \le 2w'_x(\delta)$ so for continuous functions the two moduli are equivalent.

The most basic criterion for compactness in D is the following analog of the Arzela-Ascoli theorem.

Theorem 1.2.5 *A necessary and sufficient condition for a set A to be relatively compact in the Skorohod topology is that $\sup_{x\in A}\|x\| < \infty$ and $\lim_{\delta\to 0}\sup_{x\in A} w'_x(\delta) = 0$.*

Since the space D is separable and complete an immediate consequence of this theorem is the following tightness criterion.

Theorem 1.2.6 *A necessary and sufficient condition for a sequence P_n of probability measures on D to be tight is that*

$$(i) \quad \lim_{a\to\infty}\limsup_n P_n\left[x:\ \|x\| \ge a\right] = 0,$$

and

$$(ii) \quad \lim_{\delta\to 0}\limsup_n P_n\left[x:\ w'_x(\delta) \ge \varepsilon\right] = 0 \text{ for all } \varepsilon > 0.$$

Compactness in terms of $w''_x(\delta)$. Another useful generalization of the modulus of continuity is the following modulus

$$w''_x(\delta) = \sup_{0\le u-s\le\delta}\left[\sup_{s\le t\le u}\left(\min\left[|x(u)-x(t)|, |x(t)-x(s)|\right]\right)\right].$$

This is yet another relaxation as it is not hard to see that $w''_x(\delta) \le w'_x(\delta)$. However, once again, there is no inequality in the opposite direction: for the functions

$$x_n(t) = \begin{cases} 1, & \text{for } 0 \le t < 1/n, \\ 0, & \text{for } 1/n \le t \le 1 \end{cases},$$

we have $w''_{x_n}(\delta) = 0$ while $w'_{x_n}(\delta) = 1$ for $\delta > 1/n$ because any δ-sparse partition will still contain an interval $[0, t_1)$ with $t_1 > \delta > 1/n$ where the oscillation is equal to one. This is an end-point phenomenon which also happens for the functions

$$y_n(t) = \begin{cases} 0, & \text{for } 0 \le t < 1-1/n, \\ 1, & \text{for } 1-1/n \le t \le 1 \end{cases}.$$

Nevertheless this is the only obstacle for a compactness criterion in terms of $w''_x(\delta)$ alone. The following result takes this problem into account.

Theorem 1.2.7 *A necessary and sufficient condition for a set A to have a compact closure in the Skorohod topology is that $\sup_{x\in A}\|x\| < \infty$, $\lim_{\delta\to 0}\sup_{x\in A} w''_x(\delta) = 0$ and*

$$\lim_{\delta\to 0}\sup_{x\in A}|x(\delta)-x(0)| = 0, \text{ and } \lim_{\delta\to 0}\sup_{x\in A}|x(1^-)-x(1-\delta)| = 0.$$

A direct analog of Theorem 1.2.6 is then the following.

Theorem 1.2.8 *A necessary and sufficient condition for a sequence P_n of probability measures on D to be tight is that*

$$(i) \quad \lim_{a\to\infty} \limsup_n P_n \left[x : \ \|x\| \geq a\right] = 0,$$

and

$$(ii.1) \quad \lim_{\delta\to 0} \limsup_n P_n \left[x : \ w''_x(\delta) \geq \varepsilon\right] = 0 \text{ for all } \varepsilon > 0$$

and

$$(ii.2) \quad \begin{cases} \lim_{\delta\to 0} \limsup_n P_n \left[x : \ |x(\delta) - x(0)| \geq \varepsilon\right] = 0 \\ \lim_{\delta\to 0} \limsup_n P_n \left[x : \ |x(1^-) - x(1-\delta)| \geq \varepsilon\right] = 0. \end{cases}$$

A convenient and more practical criterion for weak convergence is the following. Given a probability measure P we denote by T_P the set of all times t such that $P[J_t] = 0$ where $J_t = \{x \in D : \ x(t) \neq x(t^-)\}$ is the set of all functions that have a jump at time t. If X is a random variable on D then we write T_X for T_P where T_P is the law of X.

Theorem 1.2.9 *Suppose that the finite-dimensional distirbutions $(X^n_{t_1}, \dots, X^n_{t_k})$ of random variables X^n defined on D converge weakly as $n \to \infty$ to $(X_{t_1}, \dots, X_{t_k})$ whenever all t_i lie in T_X, and $X_1 - X_{1-\delta}$ goes weakly to zero as $\delta \to 0$. Assume also that there exists $\beta \geq 0$ and $\alpha > 1/2$ so that for all $r \leq s \leq t$ and $\lambda > 0$ we have*

$$P\left[\min\{|X^n_s - X^n_r|, |X^n_t - X^n_s|\} \geq \lambda\right] \leq \frac{C}{\lambda^{4\beta}} |F(t) - F(s)|^{2\alpha}, \tag{1.38}$$

where F is a non-decreasing continuous function on $[0, 1]$. Then X_n converge weakly to X as $n \to \infty$.

The key estimate in the proof of Theorem 1.2.9 is that (1.38) implies that there exists a constant K that depends only on C, α and β so that

$$P\left[w''_{X^n}(\delta) \geq \varepsilon\right] \leq \frac{K}{\varepsilon^{4\beta}} (F(1) - F(0))[w_F(2\delta)]^{2\alpha-1}, \tag{1.39}$$

where w_F is the modulus of continuity of the function F. This means that (1.38) ensures that condition (ii.1) of Theorem 1.2.8 holds. A useful and verifiable condition that guarantees (1.38) is that there exist $\beta > 0$, $\alpha > 1/2$ and $C > 0$ so that

$$\mathbb{E}\left\{|X^n_s - X^n_r|^{2\beta} |X^n_t - X^n_s|^{2\beta}\right\} \leq C|t - r|^{2\alpha} \tag{1.40}$$

for all n. Then we may take $F(t) = t$ and (1.39) becomes

$$P\left[w''_{X^n}(\delta) \geq \varepsilon\right] \leq \frac{K}{\varepsilon^{4\beta}} \delta^{2\alpha-1}. \tag{1.41}$$

This is why we need $\alpha > 1/2$ in (1.40). It follows that we may use (1.40) as a substitute for condition (ii.1) in Theorem 1.2.8.

In turn, the following condition is sufficient to ensure that (1.40) holds: for any $T > 0$ and $\nu > 0$ there exists a constant $C(T, \nu)$ so that for all n, and all $0 \leq s \leq t \leq u \leq T$, we have

$$\mathbb{E}\left\{|X_n(u) - X_n(t)|^2 |X_n(t) - X_n(s)|^\nu\right\} \leq C(T, \nu)(u - t)\mathbb{E}\left\{|X_n(t) - X_n(s)|^\nu\right\}. \tag{1.42}$$

Indeed, when $\nu = 0$ in (1.42) we have

$$\mathbb{E}\left\{|X_n(u) - X_n(t)|^2\right\} \leq C(T, \nu)(u - t) \text{ for all } n \text{ and all } 0 \leq t \leq u \leq T.$$

Taking $\nu = 2$ in (1.42) we get, using the above:

$$\mathbb{E}\left\{|X_n(u) - X_n(t)|^2 |X_n(t) - X_n(s)|^2\right\} \leq C(T,\nu)(u-t)\mathbb{E}\left\{|X_n(t) - X_n(s)|^2\right\} \tag{1.43}$$
$$\leq C(T,\nu)(u-t)(t-s) \leq C(T,\nu)(u-s)^2$$

and thus (1.40) indeed holds. A somehwat more general estimate than (1.42) is a reformulation in terms of the conditional expectation

$$\mathbb{E}\left\{|X_n(t) - X_n(s)|^2 \middle| \mathcal{F}_s\right\} \leq C(T)(t-s). \tag{1.44}$$

A practical advantage of working with the conditional expectation in (1.44) is that the power of $(t-s)$ on the right is equal to one, not larger than one as in (1.40).

1.3 A limit theorem for a particle in a random flow

We now return to the question of the limiting behavior of solutions of the ray equations (1.26)

$$\dot{Z}(t) = -Q(t), \quad Z(0) = 0 \tag{1.45}$$
$$\dot{Q}(t) = \frac{1}{\varepsilon} G\left(\frac{t}{\varepsilon^2}, Z(t)\right), \quad Q(0) = 0.$$

The rigorous approach to this problem lies via understanding the more general problem of the behavior of a particle in a rapidly varying in time random flow:

$$\dot{X} = \frac{1}{\varepsilon} V\left(\frac{t}{\varepsilon^2}, X\right), \quad X(0) = x, \tag{1.46}$$

with a random function V when $\varepsilon \ll 1$. This question goes back to the papers by Khasminskii [41] from the 60's with subsequent contributions by various authors: without any attempt at completeness we mention the work of Papanicolaou and Kohler [52], and Kesten and Papanicolaou [39]. We present a version of the limit theorem due to T. Komorowski [43].

Let us explain where the scaling in (1.46) comes from, apart from coinciding with that in (1.45). To see that let us start with a dynamical system

$$\frac{dY}{dT} = v_0 V\left(\frac{T}{t_0}, \frac{Y}{x_0}\right)$$

with a random time-dependent function $V(s,x)$ and introduce non-dimensional space-time variables $X = Y/x_0$, $s = T/t_0$:

$$\frac{dX}{ds} = \varepsilon V(s, X), \quad \varepsilon = \frac{v_0 t_0}{x_0}.$$

Let us now assume that $\varepsilon \ll 1$ is a small parameter – physically this means that the time it takes the particle to pass one spatial correlation length is much larger than the correlation time of the random fluctuations. Therefore, in this regime the temporal randomness of $V(s,x)$ "dominates" the spatial variations. If we now introduce a slow time t so that $t = \varepsilon^2 s$, then in the variables (t,x) the particle obeys (1.46). The limit $\varepsilon \to 0$ now corresponds to observing the particle at times much larger than the correlation time of the random fluctuations and on the spatial scale of the order of the correlation length of the medium.

The first order equation corresponding to (1.46) is

$$\frac{\partial \phi}{\partial t} + \frac{1}{\varepsilon} V\left(\frac{t}{\varepsilon^2}, x\right) \cdot \nabla \phi = 0, \quad \phi(0,x) = \phi_0(x). \tag{1.47}$$

Its solution is $\phi(T,x) = \phi_0(X(0;x;T))$, where $X(t;x,T)$ is the solution of (1.46) with the initial data prescribed at time T: $X(T;x,T) = x$.

When does one expect the trajectories of (1.46) to behave diffusively? First of all, V has to have mean zero so that the mean displacement would not be clearly biased. Second, V should "mix things around" which means that the flow should be incompressible. It helps if dynamics at "far away" points is nearly independent: this is formalized by the mixing assumption below that eliminates the memory effect. Finally, there should be no distinguished times – this requires stationarity of V in time.

The ray equations (1.45) are not quite of the form (1.46): one should consider a slightly more general dynamics with an additional slow component $F(t,x)$:

$$\frac{dX}{dt} = \frac{1}{\varepsilon}G\left(\frac{t}{\varepsilon^2}, X\right) + F(t,X), \quad X(0) = x, \tag{1.48}$$

with a function F which we will assume to be deterministic for simplicity. Then equations (1.45) are of the form (1.48) with $X = (Z,Q)$, $G = (0,V)$ and $F = (-Q,0)$.

Assumptions on the random field

Stationarity. The random field $V(t,x)$ is strictly stationary in time and space. This means that for any $t_1, t_2, \dots, t_m \in \mathbb{R}$, $x_1, \dots, x_m \in \mathbb{R}^n$, and each $h \in \mathbb{R}$ and $y \in \mathbb{R}^n$ the joint distirbution of $V(t_1+h, x+y), V(t_2+h, x+y), \dots, V(t_m+h, x+y)$ is the same as that of $V(t_1,x), V(t_2,x), \dots, V(t_m,x)$. We will denote by $R_{nm}(t,x)$ the two-point correlation tensor of $V(t,x)$:

$$R_{nm}(t,x) = \mathbb{E}\left\{V_n(s,y)V_m(t+s,y+x)\right\}. \tag{1.49}$$

Mixing. Given $C > 0$ and $\rho > 0$ let us denote by $\mathcal{V}_a^b(C,\rho)$ the σ-algebra generated by the sets of the form $\{\omega : \ V(t,x,\omega) \in A\}$ where $a \le t \le b$, $|x| \le C(1+t^\rho)$ and A is a Borel set in $\mathbb{R}^n$. We will assume that there exists $C > 0$ and $1/2 < \rho < 1$ such that for any $m \ge 0$ the mixing coefficient

$$\beta(h;C,\rho) = \sup_t \sup_{A \in \mathcal{V}_{t+h}^\infty(C,\rho), B \in \mathcal{V}_0^t(C,\rho)} \frac{|P(A \cap B) - P(A)P(B)|}{P(B)} \tag{1.50}$$

satisfies

$$h^m \beta(h;C,\rho) \le C_m \text{ for all } h \ge 0.$$

Boundedness. The random field $V(t,x)$ has three spatial derivatives and there exists a deterministic constant $C > 0$ so that with probability one we have

$$|V(t,x)| + \left|\frac{\partial V(t,x)}{\partial x_j}\right| + \left|\frac{\partial^2 V}{\partial x_i \partial x_j}\right| + \left|\frac{\partial^3 V}{\partial x_l \partial x_i \partial x_j}\right| \le C < +\infty$$

for all $1 \le i,j,l \le n$.

Incompressibility. The field V is divergence free, that is, almost surely

$$\nabla \cdot V(t,x) = \sum_{j=1}^n \frac{\partial V_j}{\partial x_j} = 0.$$

The mixing assumption is sometimes strengthened considering larger σ-algebras $\tilde{\mathcal{V}}_a^b$ generated by the sets of the form $\{\omega : \ V(t,x,\omega) \in A\}$ where $a \le t \le b$, $x \in \mathbb{R}^n$ (there is no restriction on x now) and A is a Borel set in $\mathbb{R}^n$ with the corresponding mixing coefficient

$$\tilde{\beta}(h) = \sup_t \sup_{A \in \tilde{\mathcal{V}}_{t+h}, B \in \tilde{\mathcal{V}}_0^t} \frac{|P(A \cap B) - P(A)P(B)|}{P(B)}.$$

The stronger assumption does not apply to shifts by a mean flow, that is, random fields of the form $V(t,x) = U(x - \bar{u}t)$, where $U(x)$ is a field that is mixing in space and $\bar{u}$ is a mean flow. This is an important and interesting class of random fields that we would like to include in our consideration. The small price to pay for its inclusion is the modification of the mixing condition as in (1.50).

The spatial stationarity of $V(t,x)$ is not a necessary assumption but it allows to simplify a few expressions in what follows. This can be seen already from the formal computation in Section 1.1.3. It can, however, be dropped and we adopt it here simply for convenience. On the other hand, stationarity in time is essential for the limit theorem.

The limit theorem

Let us define the diffusion matrix

$$a_{pq} = \int_0^\infty E\left\{V_q(t,0)V_p(0,0) + V_p(t,0)V_q(0,0)\right\} dt = \int_0^\infty \left[R_{pq}(t,0) + R_{qp}(t,0)\right] dt$$

and its symmetric non-negative definite square-root matrix σ: $\sigma^2 = a$. Then the following theorem holds.

Theorem 1.3.1 *Suppose that the random field $V(t,x)$ satisfies the assumptions above. Then the process $X_\varepsilon(t)$ converges weakly as $\varepsilon \to 0$ to the limit process $\bar{X}(t)$ that satisfies a stochastic differential equation*

$$d\bar{X}(t) = F(t, \bar{X}(t))dt + \sigma dW_t.$$

Here W_t is the standard Brownian motion.

The main result of [43] is actually much more general – it applies also to non-divergence free velocities. Then the large time behavior is a sum of a large (order $1/\varepsilon$) deterministic component that comes from the flow compressibility and an order one diffusive process. Komorowski also accounts for the possible small scale variations of the random field looking at equations of the form

$$\frac{dX}{dt} = \frac{1}{\varepsilon} V\left(\frac{t}{\varepsilon^2}, \frac{X(t)}{\varepsilon^\alpha}\right)$$

with $0 \le \alpha < 1$. We will not describe these generalizations in detail here. We should also mention that when $\alpha = 1$ a new regime arises – the time it takes the particle to pass one spatial correlation length is no longer much larger than the correlation time of the random fluctuations. This seriously changes the analysis.

We will present the proof of Theorem 1.3.1 under two simplifying assumptions: first, the drift $F = 0$ and second, the matrix σ is invertible. While they do not subtract any of the essential aspects of the proof, they do shorten many expressions and calculations which are sufficiently long even without them. The proof proceeds in several steps. First, we establish a mixing lemma that translates the mixing properties of the random field into a "loss-of-memroy" effect for the trajectories. Second, using the mixing lemma we establish the tightness of the family of processes $X_\varepsilon(t)$. This is done in the space D. However, as the processes $X_\varepsilon(t)$ are all continuous the limit process also has to be continuous and convergence take place in C. In the last step we identify the limit as a Brownian motion multiplied by the matrix σ by means of the martingale characterization of the Brownian motion.

The proof of tightness

The mixing lemmas

A crucial component in many proofs of this kind is some sort of a mixing lemma. It translates the mixing properties of the random field into the mixing properties of the trajectories. At the end of the day this allows us to split expectations into product of expectations and either "justify", or explain away the closure assumptions that are often made formally. In our particular problem it explains why the formal assumption that the leading order term in the asymptotic expansion (1.28) is deterministic produced the correct answer.

We set $G_0(s_1, x) = V(s_1, x)$ and

$$G_{1,j}(s_1, s_2, x) = \sum_{p=1}^{n} V_p(s_2, x) \frac{\partial V_j(s_1, x)}{\partial x_p}, \quad j = 1, \dots, n.$$

Incompressibility of $V(t, x)$ and its spatial stationarity imply that $\mathbb{E}\{G_1(s_1, s_2, x)\} = 0$. In the next lemma we drop C and ρ in the notation for the σ-algebras $\mathcal{V}_0^s(C, \rho)$.

Lemma 1.3.2 *Fix $T \geq 0$ and let $0 \leq u \leq s \leq T$. Assume that Y is a $\mathcal{V}_0^{s/\varepsilon^2}$-measurable random vector function. Then there exists $\varepsilon_0 > 0$ and a constant $C > 0$ such that for any $0 \leq u \leq s \leq s_2 \leq s_1 \leq T$ and $0 < \varepsilon < \varepsilon_0$ we have*

$$\left|\mathbb{E}\left\{V\left(\frac{s_1}{\varepsilon^2}, X_\varepsilon(u)\right) Y\left(\frac{s}{\varepsilon^2}\right)\right\}\right| \leq C\beta(s_1 - s)\mathbb{E}\left|Y\left(\frac{s}{\varepsilon^2}\right)\right|, \tag{1.51}$$

$$\left|\mathbb{E}\left\{\frac{\partial}{\partial x_k}\left[V\left(\frac{s_1}{\varepsilon^2}, X_\varepsilon(u)\right)\right] Y\left(\frac{s}{\varepsilon^2}\right)\right\}\right| \leq C\beta(s_1 - s)\mathbb{E}\left|Y\left(\frac{s}{\varepsilon^2}\right)\right|, \tag{1.52}$$

and

$$\left|\mathbb{E}\left\{G_1\left(\frac{s_1}{\varepsilon^2}, \frac{s_2}{\varepsilon^2}, X_\varepsilon(u)\right) Y\left(\frac{s}{\varepsilon^2}\right)\right\}\right| \leq C\beta^{1/2}(s_1 - s_2)\beta^{1/2}(s_2 - s)\mathbb{E}\left|Y\left(\frac{s}{\varepsilon^2}\right)\right|, \tag{1.53}$$

$$\left|\mathbb{E}\left\{\frac{\partial}{\partial x_k}\left[G_1\left(\frac{s_1}{\varepsilon^2}, \frac{s_2}{\varepsilon^2}, X_\varepsilon(u)\right)\right] Y\left(\frac{s}{\varepsilon^2}\right)\right\}\right| \leq C\beta^{1/2}(s_1 - s_2)\beta^{1/2}(s_2 - s)\mathbb{E}\left|Y\left(\frac{s}{\varepsilon^2}\right)\right|, \tag{1.54}$$

for all $1 \leq k \leq n$.

Proof. First of all, we note that for $\rho > 1/2$, $C > 1 + \sup|V(t, x)|$ and $0 < \varepsilon < \varepsilon_0(T)$ the process $X_\varepsilon(t)$, $0 \leq t \leq u \leq T$ does not leave the ball of the radius $C(1 + u^\rho/\varepsilon^{2\rho})$ centered at the origin, and hence is $\mathcal{V}_0^{u/\varepsilon^2}(C, \rho)$ -measurable:

$$|X_\varepsilon(t)| \leq \frac{1}{\varepsilon}\int_0^u \left|V\left(\frac{s}{\varepsilon^2}, X_\varepsilon(s)\right)\right| ds \leq \frac{Cu}{\varepsilon} \leq C\left(1 + \frac{u^\rho}{\varepsilon^{2\rho}}\right)$$

for all $0 \leq t \leq u$.

We first prove (1.51)-(1.52). We prove only the second inequality, (1.52) as the proof of (1.51) is identical. The idea is to replace the random variable $X_\varepsilon(u)$ by a deterministic value and use the mixing properties of the field $V(t, x)$ in time. Let $M \in \mathbb{N}$ be a fixed positive integer and $l \in \mathbb{Z}^n$. Define the event

$$A(l) = \left[\omega : \frac{l_j}{M} \leq X_j^\varepsilon(u) < \frac{l_j + 1}{M}, \quad j = 1, \dots, n\right], \quad l = (l_1, \dots, l_n).$$

The event $A(l)$ is $\mathcal{V}_0^{s/\varepsilon^2}$ measurable since $u \leq s$. When M is sufficiently large, that is, if

$$\frac{1}{M} \leq C\left(1 + \frac{t^\rho}{\varepsilon^{2\rho}}\right),$$

for almost every realization ω there exists exactly one $l \in \mathbb{Z}^n$ so that $\omega \in A(k)$. Then we may decompose the expectation in (1.52) using the fact that the random variable $X_\varepsilon(u)$ is close to the non-random value l/M on the event $A(l)$ as follows:

$$\left|\mathbb{E}\left\{\frac{\partial}{\partial x_k}\left[V\left(\frac{t}{\varepsilon^2}, X_\varepsilon(u)\right)\right] Y\left(\frac{s}{\varepsilon^2}\right)\right\}\right| = \left|\sum_l \mathbb{E}\left\{\frac{\partial}{\partial x_k}\left[V\left(\frac{t}{\varepsilon^2}, X_\varepsilon(u)\right)\right] Y\left(\frac{s}{\varepsilon^2}\right)\chi_{A(l)}\right\}\right|$$
$$\leq \left|\sum_l \mathbb{E}\left\{\left[\frac{\partial}{\partial x_k}\left[V\left(\frac{t}{\varepsilon^2}, X_\varepsilon(u)\right)\right] - \frac{\partial}{\partial x_k}\left[V\left(\frac{t}{\varepsilon^2}, \frac{l}{M}\right)\right]\right] Y\left(\frac{s}{\varepsilon^2}\right)\chi_{A(l)}\right\}\right|$$
$$+\left|\sum_l \mathbb{E}\left\{\frac{\partial}{\partial x_k}\left[V\left(\frac{t}{\varepsilon^2}, \frac{l}{M}\right)\right] Y\left(\frac{s}{\varepsilon^2}\right)\chi_{A(l)}\right\}\right| = I + II.$$

The second term above may be now estimated using the mixing property (1.50) and the fact that $\mathbb{E}\left\{\partial V/\partial x_k\right\} = 0$ by

$$II \leq 2K\beta\left(\frac{t-s}{\varepsilon^2}\right)\sum_l \mathbb{E}\left\{\left|Y\left(\frac{s}{\varepsilon^2}\right)\right|\chi_{A(l)}\right\} = 2K\beta\left(\frac{t-s}{\varepsilon^2}\right)\mathbb{E}\left\{\left|Y\left(\frac{s}{\varepsilon^2}\right)\right|\right\},$$

uniformly in M.

As we have assumed that two spatial derivatives of the field $V(t,x)$ are bounded by a deterministic constant, $\partial V/\partial x_k$ is uniformly continuous in space. Therefore, using the Lebesgue dominated convergence theorem we conclude that $I \to 0$ as $M \to +\infty$ and (1.52) follows. An identical proof shows that in addition we have the same bound for the second derivatives of the random field V:

$$\left|\mathbb{E}\left\{\frac{\partial^2}{\partial x_k \partial x_m}\left[V\left(\frac{s_1}{\varepsilon^2}, X_\varepsilon(u)\right)\right] Y\left(\frac{s}{\varepsilon^2}\right)\right\}\right| \leq C\beta(s_1 - s)\mathbb{E}\left|Y\left(\frac{s}{\varepsilon^2}\right)\right|. \tag{1.55}$$

We now prove (1.54) – the proof of (1.53) is identical. Let us first write out the expression for G_1 :

$$\left|\mathbb{E}\left\{\frac{\partial}{\partial x_k}\left[G_1\left(\frac{s_1}{\varepsilon^2}, \frac{s_2}{\varepsilon^2}, X_\varepsilon(u)\right)\right] Y\left(\frac{s}{\varepsilon^2}\right)\right\}\right|$$
$$\leq \sum_{p=1}^n \left|\mathbb{E}\left\{\frac{\partial}{\partial x_k}\left[V_p\left(\frac{s_2}{\varepsilon^2}, X_\varepsilon(u)\right)\frac{\partial}{\partial x_p}\left(V\left(\frac{s_1}{\varepsilon^2}, X_\varepsilon(u)\right)\right)\right] Y\left(\frac{s}{\varepsilon^2}\right)\right\}\right|$$

Now we may apply (1.52), (1.55) in two different ways using different parts of the inequality

$$0 \leq u \leq s \leq s_2 \leq s_1.$$

First, we may use (1.52), (1.55) with the gap between s_1 and s_2, that is, we group into "Y" in (1.52), (1.55) all terms that involve s and s_2. Using in addition the uniform bounds on V and its derivatives this leads to

$$\left|\mathbb{E}\left\{\frac{\partial}{\partial x_k}\left[G_1\left(\frac{s_1}{\varepsilon^2}, \frac{s_2}{\varepsilon^2}, X_\varepsilon(u)\right)\right] Y\left(\frac{s}{\varepsilon^2}\right)\right\}\right| \leq C\beta\left(\frac{s_1 - s_2}{\varepsilon^2}\right)\mathbb{E}\left\{\left|Y\left(\frac{s}{\varepsilon^2}\right)\right|\right\}.$$

Second, note that (1.52) may be slightly generalized to apply with $\partial V/\partial x_k$ replaced by a sufficiently smooth in space $\mathcal{V}_{s_1}^T$ random variable with an expectation equal to zero. As $\mathbb{E}\{G_1\} = 0$ indeed, we can use use this modified version of (1.52) with the gap between s_2 and s, taking "Y" in (1.52) to be simply $Y(s/\varepsilon^2)$:

$$\left|\mathbb{E}\left\{\frac{\partial}{\partial x_k}\left[G_1\left(\frac{s_1}{\varepsilon^2}, \frac{s_2}{\varepsilon^2}, X_\varepsilon(u)\right)\right] Y\left(\frac{s}{\varepsilon^2}\right)\right\}\right| \leq C\beta\left(\frac{s_2 - s}{\varepsilon^2}\right)\mathbb{E}\left\{\left|Y\left(\frac{s}{\varepsilon^2}\right)\right|\right\}.$$

Multiplying these two inequalities and taking the square root we conculde that (1.54) holds. This finishes the proof of Lemma 1.3.2. □

The proof of tightness

We will establish the inequality

$$\mathbb{E}\left\{|X_\varepsilon(t)-X_\varepsilon(s)|^2|X_\varepsilon(s)-X_\varepsilon(u)|^2\right\} \le C(t-u)^{1+\nu} \tag{1.56}$$

with $\nu > 0$. This is the criterion (1.40) for tightness in the space D. The main step in the proof is to find $\gamma \in (1,2)$ such that for all times t and s such that $t-s > 10\varepsilon^\gamma$ we have an estimate for the conditional expectation

$$\mathbb{E}\left\{|X_\varepsilon(t)-X_\varepsilon(s)|^2\Big|\,\mathcal{F}_s\right\} \le C(t-s) \text{ for } t-s > 10\varepsilon^\gamma. \tag{1.57}$$

Step 0. Nearby times. As we have explained before, the estimate (1.57) itself is sufficient to establish tightness in D for the family $X_\varepsilon(t)$ if it were to hold for all $t > s$. As we will prove it only for pairs of time with a gap: $t-s > 10\varepsilon^\gamma$, we may at the moment conclude only that

$$\mathbb{E}\left\{|X_\varepsilon(t)-X_\varepsilon(s)|^2|X_\varepsilon(s)-X_\varepsilon(u)|^2\right\} \le C(t-u)^{1+\nu} \text{ for } t-s > 10\varepsilon^\gamma \text{ and } s-u > 10\varepsilon^\gamma.$$

Our first step is to establish that, with an appropriate choice of $\gamma \in (1,2)$, if either $t-s \le 10\varepsilon^\gamma$ or $s-u \le 10\varepsilon^\gamma$, the estimate (1.56) follows from (1.57) together with the dynamical system (1.48) governing $X_\varepsilon(t)$. If both $t-s \le 10\varepsilon^\gamma$ and $s-u \le 10\varepsilon^\gamma$ then we have directly from (1.48):

$$\begin{aligned}&\mathbb{E}\left\{|X_\varepsilon(t)-X_\varepsilon(s)|^2|X_\varepsilon(s)-X_\varepsilon(u)|^2\right\} \le C\varepsilon^{-4}(t-s)^2(s-u)^2\\ &\le C\varepsilon^{11\gamma/4-4}(t-u)^{5/4} \le C(t-u)^{5/4}\end{aligned}$$

provided that $\gamma > 16/11$. On the other hand, if, say, $t-s \le 10\varepsilon^\gamma$ but $s-u > 10\varepsilon^\gamma$, (1.57) implies that

$$\mathbb{E}\left\{|X(s)-X(u)|^2\right\} \le C(s-u),$$

and (1.48) implies that with probability one

$$|X(t)-X(s)| \le \frac{C(t-s)}{\varepsilon}.$$

Therefore, the following estimate holds for such times t, s and u:

$$\begin{aligned}&\mathbb{E}\left\{|X_\varepsilon(t)-X_\varepsilon(s)|^2|X_\varepsilon(s)-X_\varepsilon(u)|^2\right\} \le \frac{C}{\varepsilon^2}(t-s)^2(s-u)\\ &\le C\varepsilon^{7\gamma/4-2}(t-u)^{5/4} \le C(t-u)^{5/4},\end{aligned}$$

provided that $\gamma > 8/7$. We see that, indeed, (1.57) together with (1.48) are sufficient to prove the tightness criterion (1.56). The rest of the proof of tightness of the processes $X_\varepsilon(t)$ is concerned with verifying (1.57).

Step 1. Taking a time-step backward. We start with a pair of times $t > s$ with a gap between them: $t-s > 10\varepsilon^\gamma$. Consider a partition of the interval $[s,t]$ into subintervals of the length

$$\Delta t = l_\varepsilon = (t-s)\left(\left[\frac{t-s}{\varepsilon^\gamma}\right]\right)^{-1},$$

where $[x]$ is the integer part of x. Then the time step l_ε is such that $\varepsilon^\gamma/2 \le l_\varepsilon \le 2\varepsilon^\gamma$ and the partition is $s = t_0 < t_1 < \cdots < t_{M+1} = t$ with a time step $\Delta t = t_{i+1} - t_i = l_\varepsilon$. The parameter $\gamma \in (1,2)$ is to be defined later. The important aspect is that $\gamma < 2$ so that Δt is much larger than the velocity correlation time ε^2. The basic idea in the proof of (1.57) is "to expand $X_\varepsilon(t) - X_\varepsilon(s)$ in a Taylor series" with a "large" time step $O(\Delta t)$. This will produce explicitly computable terms which are the first two terms in this expansion. The error terms which are nominally large are shown to be small using the mixing Lemma 1.3.2.

Dropping the subscript ε of X_ε we write for $t > s$:

$$X(t) - X(s) = \frac{1}{\varepsilon}\int_s^t V\left(\frac{u}{\varepsilon^2}, X(u)\right) du = \frac{1}{\varepsilon}\sum_{i=0}^{M}\int_{t_i}^{t_{i+1}} V\left(\frac{u}{\varepsilon^2}, X(u)\right) du \tag{1.58}$$

Therefore our task is to estimate the integral inside the summation in the right side of (1.58). In the preparation for the application of the mixing lemma the integrand on the interval $t_i \le u \le t_{i+1}$ can be rewritten as

$$\begin{aligned} V\left(\frac{u}{\varepsilon^2}, X(u)\right) &= V\left(\frac{u}{\varepsilon^2}, X(t_{i-1})\right) + \int_{t_{i-1}}^{u} \frac{d}{du_1} V\left(\frac{u}{\varepsilon^2}, X(u_1)\right) du \\ &= V\left(\frac{u}{\varepsilon^2}, X(t_{i-1})\right) + \int_{t_{i-1}}^{u} \sum_{p=1}^{n} \frac{\partial}{\partial x_p}\left[V\left(\frac{u}{\varepsilon^2}, X(u_1)\right)\right]\left(\frac{1}{\varepsilon}V_p\left(\frac{u_1}{\varepsilon^2}, X(u_1)\right)\right) du_1 \\ &= V\left(\frac{u}{\varepsilon^2}, X(t_{i-1})\right) + \frac{1}{\varepsilon}\int_{t_{i-1}}^{u} G_1\left(\frac{u}{\varepsilon^2}, \frac{u_1}{\varepsilon^2}, X(u_1)\right) du_1. \end{aligned}$$

The next step is to expand G_1 as well, also around the "one-step-backward" time t_{i-1}:

$$G_1\left(\frac{u}{\varepsilon^2}, \frac{u_1}{\varepsilon^2}, X(u_1)\right) = G_1\left(\frac{u}{\varepsilon^2}, \frac{u_1}{\varepsilon^2}, X(t_{i-1})\right) + \frac{1}{\varepsilon}\int_{t_{i-1}}^{u_1} G_2\left(\frac{u}{\varepsilon^2}, \frac{u_1}{\varepsilon^2}, \frac{u_2}{\varepsilon^2}, X(u_2)\right) du_2$$

with

$$G_2(u, u_1, u_2, x) = \sum_{q=1}^{n} \frac{\partial}{\partial x_q}\left[G_1(u, u_1, x)\right] V_q(u_2, x).$$

Putting together the above calculations we see that

$$\begin{aligned} X(t) - X(s) &= \frac{1}{\varepsilon}\sum_{i=0}^{M}\int_{t_i}^{t_{i+1}} V\left(\frac{u}{\varepsilon^2}, X(u)\right) du = \frac{1}{\varepsilon}\sum_{i=0}^{M}\int_{t_i}^{t_{i+1}} V\left(\frac{u}{\varepsilon^2}, X(t_{i-1})\right) du \\ &+ \frac{1}{\varepsilon^2}\sum_{i=0}^{M}\int_{t_i}^{t_{i+1}}\left[\int_{t_{i-1}}^{u} G_1\left(\frac{u}{\varepsilon^2}, \frac{u_1}{\varepsilon^2}, X(u_1)\right) du_1\right] du \\ &= \frac{1}{\varepsilon}\sum_{i=0}^{M}\int_{t_i}^{t_{i+1}} V\left(\frac{u}{\varepsilon^2}, X(t_{i-1})\right) du + \frac{1}{\varepsilon^2}\sum_{i=0}^{M}\int_{t_i}^{t_{i+1}}\left[\int_{t_{i-1}}^{u} G_1\left(\frac{u}{\varepsilon^2}, \frac{u_1}{\varepsilon^2}, X(t_{i-1})\right) du_1\right] du \\ &+ \frac{1}{\varepsilon^3}\sum_{i=0}^{M}\int_{t_i}^{t_{i+1}}\left[\int_{t_{i-1}}^{u}\left[\int_{t_{i-1}}^{u_1} G_2\left(\frac{u}{\varepsilon^2}, \frac{u_1}{\varepsilon^2}, \frac{u_2}{\varepsilon^2}, X(u_2)\right) du_2\right] du_1\right] du. \end{aligned}$$

The triple integral on the last line is deterministically small with an appropriate choice of γ: the time interval in each integration is smaller than ε^γ and the total number of terms is at most $2(t-s)/\varepsilon^\gamma$ as we have assumed that $t - s \ge 10\varepsilon^\gamma$. Therefore, the last integral is bounded by

$$\frac{1}{\varepsilon^3}\left|\sum_{i=0}^{M}\int_{t_i}^{t_{i+1}}\left[\int_{t_{i-1}}^{u}\left[\int_{t_{i-1}}^{u_1} G_2\left(\frac{u}{\varepsilon^2}, \frac{u_1}{\varepsilon^2}, \frac{u_2}{\varepsilon^2}, X(u_2)\right) du_2\right] du_1\right] du\right| \le C\varepsilon^{2\gamma-3}(t-s)$$

which is small if $\gamma > 3/2$. This is a general idea in proofs of weak coupling limits: pull back one time step and expand the integrands until they become almost surely small, then compute the limit of the (very) finite number of surviviing terms. In our present case we have shown that, for $3/2 < \gamma < 2$,

$$X(t) - X(s) = L_1(s,t) + L_2(s,t) + E(s,t)$$

where

$$L_1(s,t) = \frac{1}{\varepsilon}\sum_{i=0}^{M}\int_{t_i}^{t_{i+1}} V\left(\frac{u}{\varepsilon^2}, X(t_{i-1})\right) du$$

and

$$L_2(s,t) = \frac{1}{\varepsilon^2}\sum_{i=0}^{M}\int_{t_i}^{t_{i+1}} \left[\int_{t_{i-1}}^{u} G_1\left(\frac{u}{\varepsilon^2}, \frac{u_1}{\varepsilon^2}, X(t_{i-1})\right) du_1\right] du,$$

while $|E(s,t)| \le C\varepsilon^{\alpha}(t-s)$ with some $\alpha > 0$ and a deterministic constant $C > 0$. This finishes the first preliminary step in the proof of tightness.

Step 2. Application of the tightness criterion. Now we are ready to prove (1.57). That is, we have to verify that for any non-negative and $\mathcal{V}_0^{s/\varepsilon^2}$-measurable random variable Y we have for all $0 \le s \le t \le T$ such that $t \ge s + 10\varepsilon^{\gamma}$:

$$\mathbb{E}\left\{|X(t) - X(s)|^2 Y\right\} \le C(T)(t-s)\mathbb{E}\left\{Y\right\}.$$

Our estimates in Step 1 show that it is actually enough to verify that

$$\mathbb{E}\left\{(L_j(s,t))^2 Y\right\} \le C(t-s)\mathbb{E}\{Y\}, \quad j = 1, 2.$$

An estimate for L_1. We first look at the term corresponding to L_1: it is equal to

$$\mathbb{E}\left\{(L_1(s,t))^2 Y\right\} = \frac{2}{\varepsilon^2}\sum_{i<j}\sum_{p=1}^{n}\int_{t_i}^{t_{i+1}}\int_{t_j}^{t_{j+1}} \mathbb{E}\left\{V_p\left(\frac{u}{\varepsilon^2}, X(t_{i-1})\right) V_p\left(\frac{u'}{\varepsilon^2}, X(t_{j-1})\right) Y\right\} du' du$$

$$+\frac{1}{\varepsilon^2}\sum_{j}\sum_{p=1}^{n}\int_{t_j}^{t_{j+1}}\int_{t_j}^{t_{j+1}} \mathbb{E}\left\{V_p\left(\frac{u}{\varepsilon^2}, X(t_{j-1})\right) V_p\left(\frac{u'}{\varepsilon^2}, X(t_{j-1})\right) Y\right\} du' du = \sum_{i\le j} I_{ij}.$$

The idea is to use separation between t_{i-1} and t_{j-1} and apply the mixing lemma. Accordingly we look at the cases $i \le j-2$, $i = j-1$ and $i = j$ separately as the terms end up being of a different order. The terms with $i \le j-2$ may be estimated with the help of the mixing Lemma 1.3.2 using the time gap between the times u' and $t_{j-1} \ge t_{i+1} \ge u$ which is much larger than the correlation time ε^2:

$$\sum_{j=0}^{M}\sum_{i\le j-2}|I_{ij}| \le \frac{C}{\varepsilon^2}\sum_{j=0}^{M}\sum_{i\le j-2}\int_{t_i}^{t_{i+1}}\int_{t_j}^{t_{j+1}} \beta\left(\frac{u' - t_{j-1}}{\varepsilon^2}\right)\mathbb{E}\left\{Y\right\} du' du$$

$$\le \frac{C}{\varepsilon^2}\beta\left(\varepsilon^{2-\gamma}\right)(t-s)^2 E\left\{Y\right\} \le C\varepsilon^{p}(t-s)E\left\{Y\right\}$$

for any $p > 0$ since $\gamma < 2$ and $\beta(s)$ decays faster than any power of s. The term I_3 corresponding to $i = j$ can be estimated using the mixing lemma again, using the fact that t_{j-1} is smaller than both u and u':

$$\sum_{j=0}^{M} |I_{jj}| \leq \frac{C}{\varepsilon^2} \sum_{j=0}^{M} \int\limits_{t_j}^{t_{j+1}} \int\limits_{t_j}^{t_{j+1}} \mathbb{E}\left\{ V_p\left(\frac{u}{\varepsilon^2}, X(t_{j-1})\right) V_p\left(\frac{u'}{\varepsilon^2}, X(t_{j-1})\right) Y \right\} du' du \qquad (1.59)$$

$$\leq \frac{2C}{\varepsilon^2} \sum_{j \in I} \int\limits_{t_j}^{t_{j+1}} \int\limits_{u'}^{t_{j+1}} \beta\left(\frac{u-u'}{\varepsilon^2}\right) du du' E\{Y\} \leq C(t-s)E\{Y\} \int_0^\infty \beta(u) du.$$

The integral I_2 with $i = j - 1$ is estimated similarly.

A better estimate estimate for L_1. Let us now go one step further and actually identify the limit of $E\{L_{1,j}(s,t)L_{1,m}(s,t)Y\}$ with $1 \leq j, m \leq n$. The previous calculations already show that the term corresponding to the previous I_1 (but now with V_j and V_m replacing V_p and V_p) satisfies $|I_1| \leq C\varepsilon^\alpha(t-s)E\{Y\}$ with $\alpha > 0$ so we are interested only in the limit of I_2 and I_3. The term I_3 is computed as in (1.59) with the help of the mixing lemma:

$$\sum_{j \in I} |I_{jj}| = \frac{1}{\varepsilon^2} \sum_{j=0}^{M} \int\limits_{t_j}^{t_{j+1}} \int\limits_{t_j}^{t_{j+1}} \mathbb{E}\left\{ V_j\left(\frac{u}{\varepsilon^2}, X(t_{j-1})\right) V_m\left(\frac{u'}{\varepsilon^2}, X(t_{j-1})\right) Y \right\} du' du \quad (1.60)$$

$$= \frac{1}{\varepsilon^2} \sum_{j=0}^{M} \int\limits_{t_j}^{t_{j+1}} \int\limits_{t_j}^{t_{j+1}} R_{jm}\left(\frac{u-u'}{\varepsilon^2}, 0\right) du du' E\{Y\} + o(1)(t-s)E\{Y\}$$

$$= \left[\int_{-\infty}^{\infty} R_{jm}(\tau, 0) d\tau + o(1)\right](t-s)E\{Y\}.$$

Finally, I_2 corresponding to $i = j - 1$ is computed as

$$\sum_{j \in I} |I_{j-1,j}| = \frac{1}{\varepsilon^2} \sum_{j=0}^{M} \int\limits_{t_j}^{t_{j+1}} \int\limits_{t_{j-1}}^{t_j} \mathbb{E}\left\{ V_j\left(\frac{u}{\varepsilon^2}, X(t_{j-1})\right) V_m\left(\frac{u'}{\varepsilon^2}, X(t_{j-2})\right) Y \right\} du' du \qquad (1.61)$$

$$= \frac{1}{\varepsilon^2} \sum_{j \in I} \int\limits_{t_j}^{t_{j+1}} \int\limits_{t_{j-1}}^{t_j} R_{jm}\left(\frac{u-u'}{\varepsilon^2}, 0\right) du du' E\{Y\} + o(1)(t-s)E\{Y\} = o(1)(t-s)E\{Y\}.$$

because $t_{j+1} - t_j = \varepsilon^\gamma \gg \varepsilon^2$. Therefore we actually have a more precise estimate

$$\mathbb{E}\{(L_{1,j}(s,t)L_{1,m}(s,t))Y\} = \left[\int_{-\infty}^{\infty} R_{jm}(\tau, 0) d\tau + o(1)\right](t-s)E\{Y\}. \qquad (1.62)$$

An estimate for L_2. Following the above steps one also establishes the required estimate for L_2:

$$\mathbb{E}\left\{(L_2(s,t))^2 Y\right\} \leq C(t-s)E\{Y\}. \qquad (1.63)$$

There is no reason to repeat these calculations separately for L_2 except that an even stronger estimate than (1.63) holds with an appropriate choice of γ:

$$\mathbb{E}\left\{(L_2(s,t))^2 Y\right\} \leq C\varepsilon^\alpha(t-s)E\{Y\} \qquad (1.64)$$

with $\alpha > 0$. We will need (1.64) in the identification of the limit, thus we will show it now: $\mathbb{E}\left\{(L_2(s,t))^2 Y\right\}$ is equal to

$$\frac{1}{\varepsilon^4}\sum_{i,j}\int\limits_{t_i}^{t_{i+1}} du \int\limits_{t_j}^{t_{j+1}} du' \int\limits_{t_{i-1}}^{u} du_1 \int\limits_{t_{j-1}}^{u'} du_1' \mathbb{E}\left\{G_1\left(\frac{u}{\varepsilon^2},\frac{u_1}{\varepsilon^2},X(t_{i-1})\right) G_1\left(\frac{u'}{\varepsilon^2},\frac{u_1'}{\varepsilon^2},X(t_{j-1})\right) Y\right\}.$$

Once again, you split the sum above into terms with $i \le j-2$, $i = j-1$ and $i = j$: those with $i \le j-2$ add up to

$$\frac{1}{\varepsilon^4}\sum_{i\le j-2}\int\limits_{t_i}^{t_{i+1}} du \int\limits_{t_j}^{t_{j+1}} du' \int\limits_{t_{i-1}}^{u} du_1 \int\limits_{t_{j-1}}^{u'} du_1' \mathbb{E}\left\{G_1\left(\frac{u}{\varepsilon^2},\frac{u_1}{\varepsilon^2},X(t_{i-1})\right) G_1\left(\frac{u'}{\varepsilon^2},\frac{u_1'}{\varepsilon^2},X(t_{j-1})\right) Y\right\}$$
$$\le C\varepsilon^{2\gamma-4}\beta\left(\varepsilon^{\gamma-2}\right)(t-s)^2\mathbb{E}\{Y\}.$$

We used in the above estimate the mixing lemma with the gap between t_{i-1} and t_{j-1} as well as the fact that the length of each time interval is ε^γ while the total number of terms in the sum is not more than $(2(t-s)/\varepsilon^\gamma)^2$. The important difference with L_1 is that the term with $i = j$ is also small:

$$\frac{1}{\varepsilon^4}\sum_{i}\int\limits_{t_i}^{t_{i+1}} du \int\limits_{t_i}^{t_{i+1}} du' \int\limits_{t_{i-1}}^{u} du_1 \int\limits_{t_{i-1}}^{u'} du_1' \mathbb{E}\left\{G_1\left(\frac{u}{\varepsilon^2},\frac{u_1}{\varepsilon^2},X(t_{i-1})\right) G_1\left(\frac{u'}{\varepsilon^2},\frac{u_1'}{\varepsilon^2},X(t_{i-1})\right) Y\right\}$$
$$\le C\varepsilon^{3\gamma-4}(t-s)\mathbb{E}\{Y\}$$

simply because now the number of summands is bounded by $(2(t-s)/\varepsilon^\gamma)$ (without the square). This means that if we take $\gamma > 4/3$ this term is bounded by the right side of (1.64). The contribution of the terms with $i = j-1$ is estimated identically – hence (1.64) indeed holds.

Summarizing our work so far (and restoring the missing indices) we have shown that

$$\mathbb{E}\left\{(X_m(t)-X_m(s))(X_n(t)-X_n(s))Y\right\} = \left[\int_{-\infty}^{\infty} R_{mn}(\tau,0)d\tau + o(1)\right](t-s)E\{Y\} \quad (1.65)$$

for all $t > s$ with $t - s \ge 10\varepsilon^\gamma$. This, of course, implies (1.57) and hence the tightness of the family $X_\varepsilon(t)$ follows.

Identification of the limit

In order to identify the limit, using the Levy theorem (the martingale characterization of the Brownian motion) (see, for instance, Theorem 3.16 in [37]) all we have to do is verify that the limit is continuous (that we already know) and the following two conditions hold: first,

$$\lim_{\varepsilon\to 0}\mathbb{E}\left\{\left[(X_j^\varepsilon(t)-X_j^\varepsilon(s))(X_m^\varepsilon(t)-X_m^\varepsilon(s)) - a_{jm}(t-s)\right]\Psi\right\} = 0$$

for all bounded non-negative continuous functions $\Psi = \Psi(X_\varepsilon(t_1),\dots,X_\varepsilon(t_n))$ with $0 \le t_1 \le t_2 \le \dots \le t_n \le s < t \le T$. Second, we need

$$\limsup_{\varepsilon\to 0}\mathbb{E}\left\{(X_j^\varepsilon(t))^4\right\} < +\infty$$

for all $t > 0$. These conditions allow us to conlude that the limit process is a martingale. The former condition we have already verified in the previous section in the proof of tightness. The latter may be checked using very similar arguments. This finishes the proof of Theorem 1.3.1. □

Chapter 2

Random geometric optics II: the long time limit, from rays to diffusion

Here we study the long time asymptotics of rays in a weakly random medium. This problem is analyzed in the general setting of a particle in a weakly random Hamiltonian field. This chapter is based on the results of [6] and [44].

2.1 A particle in a random Hamiltonian

We have considered in Chapter 1 the asymptotic behavior of a ray in a medium with weakly random sound speed and have seen that on a short time scale the rescaled deviation of the direction of the ray from its original value becomes a diffusion process. The long time behavior of this system is an example of the analysis of the long time, large distance behavior of a particle in a weakly random time-independent Hamiltonian flow. It turns out that this limit is also described by the momentum diffusion but now, of course, without rescaling of the momentum: the particle momentum itself undergoes the Brownian motion on the energy sphere. This intuitive result has been first proved in [40] for a classical particle in dimensions higher than two, and later extended to two dimensions with the Poisson distribution of scatterers in [22], and in a general two-dimensional setting in [45]. On the other hand, the long time limit of a momentum diffusion is the standard spatial Brownian motion. Hence, a natural question arises if it is possible to obtain such a Brownian motion directly as the limiting description in the original problem of a particle in a quenched random potential. This necessitates the control of the particle behavior over times longer than those when the momentum diffusion holds. This is what we do in this chapter.

We consider a particle that moves in an isotropic weakly random Hamiltonian flow with the Hamiltonian of the form $H_\delta(x,k) = H_0(k) + \sqrt{\delta}H_1(x,k)$, $k = |k|$, and $x, k \in \mathbb{R}^d$ with $d \geq 3$:

$$\frac{dX^\delta}{dt} = \nabla_k H_\delta, \quad \frac{dK^\delta}{dt} = -\nabla_x H_\delta, \quad X^\delta(0) = 0, \quad K^\delta(0) = k_0. \tag{2.1}$$

Here $H_0(k)$ is the background Hamiltonian and $H_1(x,k)$ is a random perturbation, while the small parameter $\delta \ll 1$ measures the relative strength of random fluctuations. One expects that the effect of the random fluctuation would be of order one on the time scale of the order

$t \sim O(\delta^{-1})$. And indeed, as we have mentioned, it has been shown in [40] that, when

$$H_\delta(x,k) = \frac{k^2}{2} + \sqrt{\delta}V(x),$$

and under certain mixing assumptions on the random potential $V(x)$, the momentum process $K^\delta(t/\delta)$ converges to a diffusion process $K(t)$ on the sphere $k = k_0$ and the rescaled spatial component $\tilde{X}^\delta(t) = \delta X^\delta(t/\delta)$ converges to $X(t) = \int_0^t K(s)ds$. This is the momentum diffusion mentioned above. Another special case,

$$H_\delta(x,k) = (c_0 + \sqrt{\delta}c_1(x))|k|, \tag{2.2}$$

arises in the geometrical optics limit of wave propagation and this is the problem we are mostly interested in these notes. Here c_0 is the background sound speed, and $c_1(x)$ is a random perturbation. This case has been considered in [6], where it has been shown that, once again, $K^\delta(t/\delta)$ converges to a diffusion process $K(t)$ on the sphere $\{k = k_0\}$ while $\tilde{X}^\delta(t) = \delta X^\delta(t/\delta)$ converges to $X(t) = c_0 \int_0^t \hat{K}(s)ds$, $\hat{K}(t) := K(t)/|K(t)|$.

We show in this chapter how the momentum diffusion may be obtained and that this analysis may be pushed beyond the time of the momentum diffusion, so that under certain assumptions concerning the mixing properties of H_1 in the spatial variable there exists $\alpha_0 > 0$ such that the process $\delta^{1+\alpha}X^\delta(t/\delta^{1+2\alpha})$ converges to the standard Brownian motion in $\mathbb{R}^d$ for all $\alpha \in (0, \alpha_0)$. The main difficulty of the proof is to obtain error estimates in the convergence of $K^\delta(\cdot)$ to the momentum diffusion. The error estimates allow us to push the analysis to times much longer than δ^{-1} where the momentum diffusion converges to the standard Brownian motion. The method of the proof is a modification of the cut-off technique used in [6] and [40].

A similar question arises in the semi-classical limit of the quantum mechanics and high frequency wave propagation. The Wigner transform, or the phase space energy density of the solution of the Schrödinger equation, is approximated in a weakly random medium by the solution of a deterministic linear Boltzmann equation. This behavior is also conjectured for the acoustic and other waves in a weakly random medium [53]. As in the momentum diffusion model for a particle, the long time limit of the Boltzmann equation is the spatial diffusion equation. It has been recently shown by Erdös, Salmhofer and Yau in [24, 25] that, indeed, one may push the Erdös-Yau analysis of [23] beyond the times on which the Boltzmann equation holds and obtain the diffusive behavior of the energy density of the solutions of the Schrödinger equation in the weak coupling limit.

We also discuss in some detail the application our results to the problem of multiple scattering of the acoustic waves. Our approach is different from that of [24, 25] mentioned above: we first consider the random geometrical optics approximation of the wave phase space energy density. The rays in the phase space satisfy the Hamiltonian equations (2.1) with the Hamiltonian given by (2.2). Therefore, the aforementioned convergence result of the solutions of (2.1) to the standard Brownian motion, combined with the standard error estimates [6, 49] on the geometrical optics approximation of the Wigner distribution of the solutions of the wave equation, allows us to establish rigorously the diffusive behavior of the wave energy density. To the best of our knowledge, this is the first result of such kind for classical waves.

2.2 The main result and preliminaries

2.2.1 The notation

As we will avoid the singular point $k = 0$, where the dynamics may not be well-defined, it is convenient to introduce $\mathbb{R}^d_* := \mathbb{R}^d \setminus \{\mathbf{0}\}$ and $\mathbb{R}^{2d}_* := \mathbb{R}^d \times \mathbb{R}^d_*$. Also $\mathbb{S}^{d-1}_R(x)$ ($\mathbb{B}_R(x)$) shall stand for a sphere (open ball) in $\mathbb{R}^d$ of radius $R > 0$ centered at x. We shall drop writing either x, or R in the notation of the sphere (ball) in the particular cases when either $x = 0$, or $R = 1$. For a fixed $M > 0$ we define the spherical shell $A(M) := [k \in \mathbb{R}^d_* : M^{-1} \le |k| \le M]$ in the k-space, and $\mathcal{A}(M) := \mathbb{R}^d \times A(M)$ in the whole phase space. Given a vector $\mathbf{v} \in \mathbb{R}^d_*$ we denote by $\hat{\mathbf{v}} := \mathbf{v}/|\mathbf{v}| \in \mathbb{S}^{d-1}$ the unit vector in the direction of $\mathbf{v}$.

For any non-negative integers p, q, r, positive times $T > T_* \ge 0$ and a function $G : [T_*, T] \times \mathbb{R}^{2d}_* \to \mathbb{R}$ that has p, q and r derivatives in the respective variables we define

$$\|G\|^{[T_*,T]}_{p,q,r} := \sum \sup_{(t,x,k)\in[T_*,T]\times\mathbb{R}^{2d}} |\partial_t^\alpha \partial_x^\beta \partial_k^\gamma G(t,x,k)|. \tag{2.3}$$

The summation range covers all integers $0 \le \alpha \le p$ and all integer valued multi-indices $|\beta| \le q$ and $|\gamma| \le r$. In the special case when $T_* = 0$, $T = +\infty$ we write $\|G\|_{p,q,r} = \|G\|^{[0,+\infty)}_{p,q,r}$. We denote by $C_b^{p,q,r}([0,+\infty) \times \mathbb{R}^{2d}_*)$ the space of all functions G with $\|G\|_{p,q,r} < +\infty$. We shall also consider spaces of bounded and a suitable number of times continuously differentiable functions $C_b^{p,q}(\mathbb{R}^{2d}_*)$ and $C_b^p(\mathbb{R}^d_*)$ with the respective norms $\|\cdot\|_{p,q}$ and $\|\cdot\|_p$.

2.2.2 The background Hamiltonian

We assume that the background Hamiltonian $H_0(k)$ is isotropic, that is, it depends only on $k = |k|$, and is uniform in space. Moreover, we assume that $H_0 : [0,+\infty) \to \mathbb{R}$ is a strictly increasing function satisfying $H_0(0) \ge 0$ and such that it is of C^3-class of regularity in $(0,+\infty)$ with $H_0'(k) > 0$ for all $k > 0$, and let

$$h^*(M) := \max_{k\in[M^{-1},M]} (H_0'(k) + |H_0''(k)| + |H_0'''(k)|), \quad h_*(M) := \min_{k\in[M^{-1},M]} H_0'(k). \tag{2.4}$$

Two examples of such Hamiltonians are the quantum Hamiltonian $H_0(k) = k^2/2$ and the acoustic wave Hamiltonian $H_0(k) = c_0 k$. The qualitative reason for the above assumptions on $H_0(k)$ is that we need the background dynamics 'to take the particle to various regions where it will sample the nearly independent random fluctuations. The overall effect will then lead to a Markovian limit. This makes the problem much simpler than a seemingly similar problem

$$\dot{X} = V(X), \tag{2.5}$$

with a mixing in space and time-independent random field $V(x)$. Unlike our problem, (2.5) lack any mechanism to move the particle around which makes it extremely difficult to obtain any rigorous, or even formal results for the particle behavior in (2.5).

2.2.3 The random medium

Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, and let $\mathbb{E}$ denote the expectation with respect to $\mathbb{P}$. We denote by $\|X\|_{L^p(\Omega)}$ the L^p-norm of a given random variable $X : \Omega \to \mathbb{R}$, $p \in [1,+\infty]$. Let $H_1 : \mathbb{R}^d \times [0,+\infty) \times \Omega \to \mathbb{R}$ be a random field that is measurable and strictly stationary in the first variable. This means that for any shift $x \in \mathbb{R}^d$, $k \in [0,+\infty)$, and a collection of points $x_1, \ldots, x_n \in \mathbb{R}^d$ the laws of $(H_1(x_1 + x, k), \ldots, H_1(x_n + x, k))$ and $(H_1(x_1, k), \ldots, H_1(x_n, k))$ are identical. In addition, we assume that $\mathbb{E}H_1(x,k) = 0$ for all $k \ge 0$, $x \in \mathbb{R}^d$, the realizations of $H_1(x,k)$ are $\mathbb{P}$–a.s. C^2-smooth in $(x,k) \in \mathbb{R}^d \times (0,+\infty)$ and they satisfy

$$D_{i,j}(M) := \max_{|\alpha|=i} \operatorname*{ess-sup}_{(x,k,\omega)\in\mathbb{R}^d\times[M^{-1},M]\times\Omega} |\partial_{\mathbf{x}}^\alpha \partial_k^j H_1(\mathbf{x},k;\omega)| < +\infty, \quad i,j = 0,1,2. \tag{2.6}$$

We define $\tilde{D}(M) := \sum_{0\le i+j\le 2} D_{i,j}(M)$.

We suppose further that the random field is strongly mixing in the uniform sense. More precisely, for any $R>0$ we let $\mathcal{C}_R^i$ and $\mathcal{C}_R^e$ be the σ–algebras generated by random variables $H_1(\mathbf{x},k)$ with $k\in[0,+\infty)$, $\mathbf{x}\in\mathbb{B}_R$ and $\mathbf{x}\in\mathbb{B}_R^c$ respectively. The uniform mixing coefficient between the σ–algebras is

$$\phi(\rho) := \sup[\,|\mathbb{P}(B)-\mathbb{P}(B|A)|:\ R>0,\ A\in\mathcal{C}_R^i,\ B\in\mathcal{C}_{R+\rho}^e],$$

for all $\rho>0$. We suppose that $\phi(\rho)$ decays faster than any power: for each $p>0$

$$h_p := \sup_{\rho\ge 0}\rho^p\phi(\rho)<+\infty. \tag{2.7}$$

The two-point spatial correlation function of the random field H_1 is

$$R(\mathbf{y},k) := \mathbb{E}[H_1(\mathbf{y},k)H_1(\mathbf{0},k)].$$

Note that (2.7) implies that for each $p>0$

$$h_p(M) := \sum_{i=0}^{4}\sum_{|\alpha|=i}\sup_{(\mathbf{y},k)\in\mathbb{R}^d\times[M^{-1},M]}(1+|\mathbf{y}|^2)^{p/2}|\partial_{\mathbf{y}}^\alpha R(\mathbf{y},k)|<+\infty,\qquad M>0. \tag{2.8}$$

We also assume that the correlation function $R(y,l)$ is of the C^∞-class for a fixed $l>0$, is sufficiently smooth in l, and that for any fixed $l>0$

$$\hat{R}(k,l)\ \text{does not vanish identically on any hyperplane}\ H_p=\{k:\ (k\cdot p)=0\}. \tag{2.9}$$

Here $\hat{R}(k,l)=\int R(x,l)\exp(-ik\cdot x)dx$ is the power spectrum of H_1.

The above assumptions are satisfied, for example, if $H_1(x,k)=c_1(x)h(k)$, where $c_1(x)$ is a stationary uniformly mixing random field with a smooth correlation function, and $h(k)$ is a smooth deterministic function.

2.2.4 The path-spaces

For fixed integers $d,m\ge 1$ we let $\mathcal{C}^{d,m}:=C([0,+\infty);\mathbb{R}^d\times\mathbb{R}_*^m)$: we shall omit the subscripts in the notation of the path space if $m=d$. We define $(X(t),K(t)):\mathcal{C}^{d,m}\to\mathbb{R}^d\times\mathbb{R}_*^m$ as the canonical mapping $(X(t;\pi),K(t;\pi)):=\pi(t)$, $\pi\in\mathcal{C}^{d,m}$ and also let $\theta_s(\pi)(\cdot):=\pi(\cdot+s)$ be the standard shift transformation.

For any $u\le v$ denote by $\mathcal{M}_u^v$ the σ-algebra of subsets of $\mathcal{C}$ generated by $(X(t),K(t))$, $t\in[u,v]$. We write $\mathcal{M}^v:=\mathcal{M}_0^v$ and $\mathcal{M}$ for the σ algebra of Borel subsets of $\mathcal{C}$. It coincides with the smallest σ–algebra that contains all $\mathcal{M}^t$, $t\ge 0$.

Let $\delta_*(M):=H_0\left(M^{-1}\right)/(2\tilde{D}(M))$. For a given $M>0$ and $\delta\in(0,\delta_*(M)]$ we let

$$M_\delta := \max\left\{H_0^{-1}(H_0(M)+2\sqrt{\delta}\tilde{D}(M)),\left[H_0^{-1}\left(H_0\left(\frac{1}{M}\right)-2\sqrt{\delta}\tilde{D}(M)\right)\right]^{-1}\right\}. \tag{2.10}$$

For a particle that is governed by the Hamiltonian flow generated by $H_\delta(x,k)$ we have

$$M_\delta^{-1}\le|K(t)|\le M_\delta$$

for all t provided that $K(0)\in A(M)$. Accordingly, we define $\mathcal{C}(T,\delta)$ as the set of paths $\pi\in\mathcal{C}$ so that both $(2M_\delta)^{-1}\le|K(t)|\le 2M_\delta$, and

$$\left|X(t)-X(u)-\int\limits_u^t H_0'(K(s))\hat{K}(s)ds\right|\le\tilde{D}(2M_\delta)\sqrt{\delta}(t-u),\ \text{ for all }\ 0\le u<t\le T.$$

In the case when $\delta=1$, or $T=+\infty$ we shall write simply $\mathcal{C}(T)$, or $\mathcal{C}(\delta)$ respectively.

2.2.5 The main results

Let the function $\phi_\delta(t,x,k)$ satisfy the Liouville equation

$$\begin{aligned}&\frac{\partial\phi^\delta}{\partial t}+\nabla_x H_\delta(x,k)\cdot\nabla_k\phi^\delta-\nabla_k H_\delta(x,k)\cdot\nabla_x\phi^\delta=0,\\&\phi^\delta(0,x,k)=\phi_0(\delta x,k).\end{aligned}\tag{2.11}$$

We assume that the initial data $\phi_0(x,k)$ is a compactly supported function four times differentiable in k, twice differentiable in x whose support is contained inside a spherical shell $\mathcal{A}(M)=\{(x,k):M^{-1}<|k|<M\}$ for some positive $M>0$.

Let us define the diffusion matrix D_{mn} by

$$D_{mn}(\hat{k},l)=-\frac{1}{2}\int_{-\infty}^{\infty}\frac{\partial^2 R(H_0'(l)s\hat{k},l)}{\partial x_n\partial x_m}ds=-\frac{1}{2H_0'(l)}\int_{-\infty}^{\infty}\frac{\partial^2 R(s\hat{k},l)}{\partial x_n\partial x_m}ds,\quad m,n=1,\ldots,d.\tag{2.12}$$

Then we have the following result.

Theorem 2.2.1 *Let ϕ^δ be the solution of* (2.11) *and let $\bar\phi$ satisfy*

$$\begin{aligned}&\frac{\partial\bar\phi}{\partial t}=\sum_{m,n=1}^{d}\frac{\partial}{\partial k_m}\left(D_{mn}(\hat{k},k)\frac{\partial\bar\phi}{\partial k_n}\right)+H_0'(k)\,\hat{k}\cdot\nabla_x\bar\phi\\&\bar\phi(0,x,k)=\phi_0(x,k).\end{aligned}\tag{2.13}$$

Suppose that $M\geq M_0>0$ and $T\geq T_0>0$. Then, there exist two constants C, $\alpha_0>0$ such that for all $T\geq T_0$

$$\sup_{(t,x,k)\in[0,T]\times K}\left|\mathbb{E}\phi^\delta\left(\frac{t}{\delta},\frac{x}{\delta},k\right)-\bar\phi(t,x,k)\right|\leq CT(1+\|\phi_0\|_{1,4})\delta^{\alpha_0}\tag{2.14}$$

for all compact sets $K\subset\mathcal{A}(M)$.

Remark 2.2.2 We shall denote by C, $C_1,\ldots$, α_0, $\alpha_1,\ldots$, γ_0, $\gamma_1,\ldots$ throughout this article generic positive constants. Unless specified otherwise the constants denoted this way *shall depend neither on δ, nor on T*. We will also assume that $T\geq T_0>0$ and $M\geq M_0>0$.

Remark 2.2.3 Classical results of the theory of stochastic differential equations, see e.g. Theorem 6 of Chapter 2, p. 176 and Corollary 4 of Chapter 3, p. 303 of [32], imply that there exists a unique solution to the Cauchy problem (2.13) that belongs to the class $C_b^{1,1,2}([0,+\infty)\times\mathbb{R}^{2d}_*)$. This solution admits a probabilistic representation using the law of a time homogeneous diffusion $\mathfrak{Q}_{x,k}$ whose Kolmogorov equation is given by (2.13), see Section 2.3.3 below.

Note that

$$\begin{aligned}\sum_{m=1}^{d}D_{nm}(\hat{k},k)\hat{k}_m&=-\sum_{m=1}^{d}\frac{1}{2H_0'(k)}\int_{-\infty}^{\infty}\frac{\partial^2 R(s\hat{k},k)}{\partial x_n\partial x_m}\hat{k}_m ds\\&=-\sum_{m=1}^{d}\frac{1}{2H_0'(k)}\int_{-\infty}^{\infty}\frac{d}{ds}\left(\frac{\partial R(s\hat{k},k)}{\partial x_n}\right)ds=0\end{aligned}$$

and thus the K-process generated by (2.13) is indeed a diffusion process on a sphere $k=\text{const}$, or, equivalently, equations (2.13) for different values of k are decoupled. Assumption (2.9) implies the following.

Proposition 2.2.4 *The matrix $D(\hat{k}, l)$ has rank $d-1$ for each $\hat{k} \in \mathbb{S}^{d-1}$ and each $l > 0$.*

The proof can be found in [6] (Proposition 4.3 in [6]). It can be shown, using the argument given on pp. 122-123 of this paper that, under assumption (2.9), equation (2.13) is hypoelliptic on the manifold $\mathbb{R}^d \times \mathbb{S}_k^{d-1}$ for each $k > 0$.

We also show that solutions of (2.13) converge in the long time limit to the solutions of the spatial diffusion equation. More, precisely, we have the following result. Let $\bar{\phi}_\gamma(t, x, k) = \bar{\phi}(t/\gamma^2, x/\gamma, k)$, where $\bar{\phi}$ satisfies (2.13) with an initial data $\bar{\phi}_\gamma(0, t, x, k) = \phi_0(\gamma x, k)$. We also let $w(t, x, k)$ be the solution of the spatial diffusion equation:

$$\frac{\partial w}{\partial t} = \sum_{m,n=1}^{d} a_{mn}(k) \frac{\partial^2 w}{\partial x_n \partial x_m}, \tag{2.15}$$
$$w(0, x, k) = \bar{\phi}_0(x, k)$$

with the averaged initial data

$$\bar{\phi}_0(x, k) = \frac{1}{\Gamma_{d-1}} \int_{\mathbb{S}^{d-1}} \phi_0(x, k) d\Omega(\hat{k}).$$

Here $d\Omega(\hat{k})$ is the surface measure on the unit sphere $\mathbb{S}^{d-1}$ and Γ_n is the area of an n-dimensional sphere. The diffusion matrix $A := [a_{nm}]$ in (2.15) is given explicitly as

$$a_{nm}(k) = \frac{1}{\Gamma_{d-1}} \int_{\mathbb{S}^{d-1}} H_0'(k) \hat{k}_n \chi_m(k) d\Omega(\hat{k}). \tag{2.16}$$

The functions χ_j appearing above are the mean-zero solutions of

$$\sum_{m,n=1}^{d} \frac{\partial}{\partial k_m} \left(D_{mn}(\hat{k}, k) \frac{\partial \chi_j}{\partial k_n} \right) = -H_0'(k) \hat{k}_j. \tag{2.17}$$

Note that equations (2.17) for χ_m are elliptic on each sphere $\{|k| = k\}$. This follows from the fact that the equations for each such sphere are all decoupled and Proposition 2.2.4. Also note that the matrix A is positive definite. Indeed, let $\mathbf{c} = (c_1, \ldots, c_d) \in \mathbb{R}^d$ be a fixed vector and let $\chi_{\mathbf{c}} := \sum_{m=1}^{d} c_m \chi_m$. Since the matrix D is non-negative we have

$$\begin{aligned}
(A\mathbf{c}, \mathbf{c})_{\mathbb{R}^d} &= -\frac{1}{\Gamma_{d-1}} \sum_{m,n=1}^{d} \int_{\mathbb{S}^{d-1}} \chi_{\mathbf{c}}(\hat{k}, l) \frac{\partial}{\partial k_m} \left(D_{mn}(\hat{k}, l) \frac{\partial \chi_{\mathbf{c}}(\hat{k}, l)}{\partial k_n} \right) d\Omega(\hat{k}) \\
&= -\frac{1}{\Gamma_{d-1}} \sum_{m,n=1}^{d} \int_{\mathbb{R}^d} \chi_{\mathbf{c}}(\hat{k}, l) \frac{\partial}{\partial k_m} \left(D_{mn}(\hat{k}, l) \frac{\partial \chi_{\mathbf{c}}(\hat{k}, l)}{\partial k_n} \right) \delta(k - l) \frac{dk}{l^{d-1}} \\
&= \frac{1}{\Gamma_{d-1}} \int_{\mathbb{S}^{d-1}} (D(\hat{k}, l) \nabla \chi_{\mathbf{c}}(\hat{k}, l), \nabla \chi_{\mathbf{c}}(\hat{k}, l))_{\mathbb{R}^d} d\hat{\Omega}(\hat{k}) \geq 0.
\end{aligned} \tag{2.18}$$

The last equality holds after integration by parts because $D(\hat{k}, l)\hat{k} = 0$. Moreover, the inequality appearing in the last line of (2.18) is strict. This can be seen as follows. Since the null-space of the matrix $D(\hat{k}, l)$ is one-dimensional and consists of the vectors parallel to $\hat{k}$, in order for $(A\mathbf{c}, \mathbf{c})_{\mathbb{R}^d}$ to vanish one needs that the gradient $\nabla \chi_{\mathbf{c}}(\hat{k}, l)$ is parallel to $\hat{k}$ for all $\hat{k} \in \mathbb{S}^{d-1}$. This, however, together with (2.17) would imply that $\hat{k} \cdot \mathbf{c} = 0$ for all $\hat{k}$, which is impossible.

The following theorem holds.

Theorem 2.2.5 *For every pair of times* $0 < T_* < T < +\infty$ *the re-scaled solution* $\bar\phi_\gamma(t,x,k) = \bar\phi(t/\gamma^2, x/\gamma, k)$ *of* (2.13) *converges as* $\gamma \to 0$ *in* $C([T_*,T]; L^\infty(\mathbb{R}^{2d}))$ *to* $w(t,x,k)$. *Moreover, there exists a constant* $C > 0$ *so that we have*

$$\|w(t,\cdot) - \bar\phi_\gamma(t,\cdot)\|_{0,0} \le C\left(\gamma T + \sqrt{\gamma}\right)\|\phi_0\|_{1,1} \tag{2.19}$$

for all $T_* \le t \le T$.

Remark 2.2.6 In fact, as it will become apparent in the course of the proof, we have a stronger result, namely T_* can be made to vanish as $\gamma \to 0$. For instance, we can choose $T_* = \gamma^{3/2}$, see (2.105). We can not set $T_* = 0$ since there is a small initial layer when the solution of (2.13) adjusts to become independent of the direction $\hat k$.

The proof of Theorem 2.2.5 is based on some classical asymptotic expansions and is quite straightforward. As an immediate corollary of Theorems 2.2.1 and 2.2.5 we obtain the following result, which is the main result of this chapter.

Theorem 2.2.7 *Let* ϕ_δ *be solution of* (2.11) *with the initial data* $\phi_\delta(0,x,k) = \phi_0(\delta^{1+\alpha}x, k)$ *and let* $\bar w(t,x)$ *be the solution of the diffusion equation* (2.15) *with the initial data* $w(0,x,k) = \bar\phi_0(x,k)$. *Then, there exists* $\alpha_0 > 0$ *and a constant* $C > 0$ *so that for all* $0 \le \alpha < \alpha_0$ *and all* $0 < T_* \le T$ *we have for all compact sets* $K \subset \mathcal{A}(M)$:

$$\sup_{(t,x,k)\in[T_*,T]\times K} \left|w(t,x,k) - \mathbb{E}\bar\phi_\delta(t,x,k)\right| \le CT\delta^{\alpha_0-\alpha}, \tag{2.20}$$

where $\bar\phi_\delta(t,x,k) := \phi_\delta\left(t/\delta^{1+2\alpha}, x/\delta^{1+\alpha}, k\right)$.

Theorem 2.2.7 shows that the movement of a particle in a weakly random quenched Hamiltonian is, indeed, approximated by a Brownian motion in the long time-large space limit, at least for times $T \ll \delta^{-\alpha_0}$. In fact, according to Remark 2.2.6 we can allow T_* to vanish as $\delta \to 0$ choosing $T_* = \delta^{3\alpha/2}$.

In the isotropic case when $R = R(|x|, k)$ we may simplify the above expressions for the diffusion matrices D_{mn} and a_{mn}. In that case we have

$$\begin{aligned} D_{mn}(\hat k, k) &= -\frac{1}{2}\int_{-\infty}^{\infty} \frac{\partial^2 R(H_0'(k)s\hat k, k)}{\partial x_n \partial x_m}\,ds \\ &= -\int_0^\infty \left[\frac{k_n k_m}{k^2} R''(H_0'(k)s, k) + \left(\delta_{nm} - \frac{k_n k_m}{k^2}\right)\frac{R'(H_0'(k)s,k)}{H_0'(k)s}\right] ds \\ &= -\frac{1}{H_0'(k)}\int_0^\infty \frac{R'(s,k)}{s}\,ds\,\left(\delta_{nm} - \frac{k_n k_m}{k^2}\right), \end{aligned}$$

so that the matrix $[D_{mn}(\hat k, k)]$ has the form

$$D(\hat k, k) = D_0(k)\left(I - \hat k \otimes \hat k\right), \quad D_0(k) = -\frac{1}{H_0'(k)}\int_0^\infty \frac{R'(s,k)}{s} ds.$$

The functions χ_j are given explicitly in this case by

$$\chi_j(\hat k, k) = -\frac{|H_0'(k)|^2 |k|^2 \hat k_j}{(d-1)\bar D_0(k)}, \quad \bar D_0(k) = -\int_0^\infty \frac{R'(s,k)}{s} ds$$

and

$$a_{nm}(k) = \frac{|H_0'(k)|^3 |k|^2}{\Gamma_{d-1}(d-1)\bar D_0(k)}\int_{\mathbb{S}^{d-1}} \hat k_n \hat k_m d\Omega(\hat k) = \frac{|H_0'(k)|^3 |k|^2}{d(d-1)\bar D_0(k)}\delta_{nm}.$$

2.2.6 A formal derivation of the momentum diffusion

We now recall how the diffusion operator in (2.13) can be derived in a quick formal way. We represent the solution of (2.11) as $\phi^\delta(t,x,k) = \psi^\delta(\delta t, \delta x, k)$ and write an asymptotic multiple scale expansion for ψ^δ

$$\psi^\delta(t,x,k) = \bar{\phi}(t,x,k) + \sqrt{\delta}\phi_1\left(t,x,\frac{x}{\delta},k\right) + \delta\phi_2\left(t,x,\frac{x}{\delta},k\right) + \dots \tag{2.21}$$

We assume formally that the leading order term $\bar{\phi}$ is deterministic and independent of the fast variable $z = x/\delta$. We insert this expansion into (2.11) and obtain in the order $O\left(\delta^{-1/2}\right)$:

$$\nabla_z H_1(z,k)\cdot\nabla_k\bar{\phi} - H_0'(k)\hat{k}\cdot\nabla_z\phi_1 = 0. \tag{2.22}$$

Let $\theta \ll 1$ be a small positive regularization parameter that will be later sent to zero, and consider a regularized version of (2.22):

$$\frac{1}{H_0'(k)}\nabla_z H_1(z,k)\cdot\nabla_k\bar{\phi} - \hat{k}\cdot\nabla_z\phi_1 + \theta\phi_1 = 0,$$

Its solution is

$$\phi_1(z,k) = -\frac{1}{H_0'(k)}\int_0^\infty \sum_{m=1}^d \frac{\partial H_1(z+s\hat{k},k)}{\partial z_m}\frac{\partial\bar{\phi}(t,x,k)}{\partial k_m}e^{-\theta s}ds. \tag{2.23}$$

The next order equation becomes upon averaging

$$\frac{\partial\bar{\phi}}{\partial t} = \mathbb{E}\left(\frac{\partial H_1(z,k)}{\partial k}\hat{k}\cdot\nabla_z\phi_1\right) - \mathbb{E}\left(\nabla_z H_1(z,k)\cdot\nabla_k\phi_1\right) + H_0'(k)\hat{k}\cdot\nabla_x\bar{\phi}. \tag{2.24}$$

The first two terms on the right hand side above may be computed explicitly using expression (2.23) for ϕ_1:

$$\begin{aligned}
&\mathbb{E}\left(\frac{\partial H_1(z,k)}{\partial k}\hat{k}\cdot\nabla_z\phi_1\right) - \mathbb{E}\left(\nabla_z H_1(z,k)\cdot\nabla_k\phi_1\right)\\
&= -\mathbb{E}\left[\sum_{m,n=1}^d \frac{\partial H_1(z,k)}{\partial k}\hat{k}_m\frac{\partial}{\partial z_m}\left(\frac{1}{H_0'(k)}\int_0^\infty \frac{\partial H_1(z+s\hat{k},k)}{\partial z_n}\frac{\partial\bar{\phi}(t,x,k)}{\partial k_n}e^{-\theta s}ds\right)\right]\\
&+\mathbb{E}\left[\sum_{m,n=1}^d \frac{\partial H_1(z,k)}{\partial z_m}\frac{\partial}{\partial k_m}\left(\frac{1}{H_0'(k)}\int_0^\infty \frac{\partial H_1(z+s\hat{k},k)}{\partial z_n}\frac{\partial\bar{\phi}(t,x,k)}{\partial k_n}e^{-\theta s}ds\right)\right].
\end{aligned}$$

Using spatial stationarity of $H_1(z,k)$ we may rewrite the above as

$$
\begin{aligned}
&-\mathbb{E}\left[\sum_{m,n=1}^{d}\frac{\partial H_1(z,k)}{\partial k}\hat{k}_m\frac{\partial}{\partial z_m}\left(\frac{1}{H_0'(k)}\int_0^{\infty}\frac{\partial H_1(z+s\hat{k},k)}{\partial z_n}\frac{\partial\bar{\phi}(t,x,k)}{\partial k_n}e^{-\theta s}ds\right)\right]\\
&-\mathbb{E}\left[\sum_{m,n=1}^{d}H_1(z,k)\frac{\partial}{\partial z_m}\frac{\partial}{\partial k_m}\left(\frac{1}{H_0'(k)}\int_0^{\infty}\frac{\partial H_1(z+s\hat{k},k)}{\partial z_n}\frac{\partial\bar{\phi}(t,x,k)}{\partial k_n}e^{-\theta s}ds\right)\right]\\
&=-\sum_{m,n=1}^{d}\frac{\partial}{\partial k_m}\left[\frac{1}{H_0'(k)}\int_0^{\infty}\mathbb{E}\left(H_1(z,k)\frac{\partial^2 H_1(z+s\hat{k},k)}{\partial z_n\partial z_m}\right)\frac{\partial\bar{\phi}(t,x,k)}{\partial k_n}e^{-\theta s}ds\right]\\
&=-\sum_{m,n=1}^{d}\frac{\partial}{\partial k_m}\left(\frac{1}{H_0'(k)}\int_0^{\infty}\frac{\partial^2 R(s\hat{k},k)}{\partial x_n\partial x_m}\frac{\partial\bar{\phi}(t,x,k)}{\partial k_n}e^{-\theta s}ds\right)\\
&\to-\frac{1}{2}\sum_{m,n=1}^{d}\frac{\partial}{\partial k_m}\left(\frac{1}{H_0'(k)}\int_{-\infty}^{\infty}\frac{\partial^2 R(s\hat{k},k)}{\partial x_n\partial x_m}\frac{\partial\bar{\phi}(t,x,k)}{\partial k_n}ds\right),\ \text{as }\theta\to 0^+.
\end{aligned}
$$

We insert the above expression into (2.24) and obtain

$$
\frac{\partial\bar{\phi}}{\partial t}=\sum_{m,n=1}^{d}\frac{\partial}{\partial k_n}\left(D_{nm}(\hat{k},k)\frac{\partial\bar{\phi}}{\partial k_m}\right)+H_0'(k)\hat{k}\cdot\nabla_x\bar{\phi} \tag{2.25}
$$

with the diffusion matrix $D(\hat{k},k)$ as in (2.12). Observe that (2.25) is nothing but (2.13). However, the naive asymptotic expansion (2.21) may not be justified. The rigorous proof presented in the next section is based on a quite different method.

2.3 From the Liouville equation to the momentum diffusion. Estimation of the convergence rates: proof of Theorem 2.2.1

Outline of the proof

The basic idea of the proof of Theorem 2.2.1 is a modification of that of [6, 40]. We consider the trajectories corresponding to the Liouville equation (2.11) and introduce a stopping time, called τ_δ, that, among others, prevents near self-intersection of trajectories. This fact ensures that until the stopping time occurs the particle is "exploring a new territory" and, thanks to the strong mixing properties of the medium, "memory effects" are lost. Therefore, roughly speaking, until the stopping time the process is approximately characterized by the Markov property. Furthermore, since the amplitude of the random Hamiltonian is not strong enough to destroy the continuity of its path, it becomes a diffusion in the limit, as $\delta\to 0$. We introduce also an augmented process that follows the trajectories of the Hamiltonian flow until the stopping time τ_δ and becomes a diffusion after $t=\tau_\delta$. We show that the law of the augmented process is close to the law of a diffusion, see Proposition 2.3.4, with an explicit error bound. We also prove that the stopping time tends to infinity as $\delta\to 0$, once again with the error bound that is proved in Theorem 2.3.6. The combination of these two results allows us to estimate the difference between the solutions of the Liouville and the diffusion equations in a rather straightforward manner (see Section 2.3.6): they are close until the stopping time as the law of the diffusion is always close to that of the augmented process, while the latter coincides with the true process until τ_δ. On the other hand, the fact that $\tau_\delta \to \infty$ as $\delta \to 0$ shows that with a large probability the augmented process is close to the true process. This combination finishes the proof.

The random characteristics corresponding to (2.11)

Consider the motion of a particle governed by a Hamiltonian system of equations

$$\begin{cases} \frac{dz^{(\delta)}(t;\mathbf{x},\mathbf{k})}{dt} = (\nabla_{\mathbf{k}} H_\delta)\left(\frac{z^{(\delta)}(t;\mathbf{x},\mathbf{k})}{\delta}, \mathbf{m}^{(\delta)}(t;\mathbf{x},\mathbf{k})\right) \\ \frac{d\mathbf{m}^{(\delta)}(t;\mathbf{x},\mathbf{k})}{dt} = -\frac{1}{\sqrt{\delta}}(\nabla_z H_\delta)\left(\frac{z^{(\delta)}(t;\mathbf{x},\mathbf{k})}{\delta}, \mathbf{m}^{(\delta)}(t;\mathbf{x},\mathbf{k})\right) \\ z^{(\delta)}(0;\mathbf{x},\mathbf{k}) = \mathbf{x}, \quad \mathbf{m}^{(\delta)}(0;\mathbf{x},\mathbf{k}) = \mathbf{k}, \end{cases} \tag{2.26}$$

where the Hamiltonian $H_\delta(x,k) := H_0(k) + \sqrt{\delta} H_1(x,k)$, $k = |k|$. The trajectories of (2.26) are the characteristics of the Liouville equation (2.11). The hypotheses made in Section 2.2 imply that the trajectory $(z^{(\delta)}(t;\mathbf{x},k), \mathbf{m}^{(\delta)}(t;\mathbf{x},k))$ necessarily lies in $\mathcal{C}(T,\delta)$ for each $T > 0$, $\delta \in (0, \delta_*(M)]$, provided that the initial data $(\mathbf{x},k) \in \mathcal{A}(M)$. Indeed, it follows from the Hamiltonian structure of (2.26) that the Hamiltonian $H_\delta(x,m) = H_0(m) + \sqrt{\delta} H_1(z,m)$ must be conserved along the trajectory. Hence, the definition (2.10) implies that $M_\delta^{-1} \le |\mathbf{m}^{(\delta)}(\cdot;\mathbf{x},k)| \le M_\delta$. We denote by $Q^\delta_{s,\mathbf{x},\mathbf{k}}(\cdot)$ the law over $\mathcal{C}$ of the process corresponding to (2.26) starting at $t = s$ from $(\mathbf{x},\mathbf{k})$ (this law is actually supported in $\mathcal{C}(\delta)$). We shall omit writing the subscript s when it equals to 0.

The stopping times

We now define the stopping time τ_δ, described in Section 2.3, that prevents the trajectories of (2.26) to have near self-intersections (recall that the intent of the stopping time is to prevent any "memory effects" of the trajectories). As we have already mentioned, we will later show that the probability of the event $[\,\tau_\delta < T\,]$ for a fixed $T > 0$ goes to zero, as $\delta \to 0$.

Let $0 < \epsilon_1 < \epsilon_2 < 1/2$, $\epsilon_3 \in (0, 1/2 - \epsilon_2)$, $\epsilon_4 \in (1/2, 1 - \epsilon_1 - \epsilon_2)$ be small positive constants that will be further determined later and set

$$N = [\delta^{-\epsilon_1}], \quad p = [\delta^{-\epsilon_2}], \quad q = p\,[\delta^{-\epsilon_3}], \quad N_1 = Np\,[\delta^{-\epsilon_4}]. \tag{2.27}$$

We will specify additional restrictions on the constants ϵ_j as the need for such constraints arises. However, the basic requirement is that ϵ_i, $i = 1,2,3$ should be sufficiently small and ϵ_4 is bigger than 1/2, less than one and can be made as close to one as we would need it. It is important that $\epsilon_1 < \epsilon_2$ so that $N \ll p$ when $\delta \ll 1$. We introduce the following $(\mathcal{M}^t)_{t\ge 0}-$stopping times. Let $t_k^{(p)} := kp^{-1}$ be a mesh of times, and $\pi \in \mathcal{C}$ be a path. We define the "violent turn" stopping time

$$S_\delta(\pi) := \inf\Bigg[t \ge 0 : \text{ for some } k \ge 0 \text{ we have } t \in \left[t_k^{(p)}, t_{k+1}^{(p)}\right) \text{ and } \tag{2.28}$$
$$\hat{K}(t_{k-1}^{(p)}) \cdot \hat{K}(t) \le 1 - \frac{1}{N}, \text{ or } \hat{K}\left(t_k^{(p)} - \frac{1}{N_1}\right) \cdot \hat{K}(t) \le 1 - \frac{1}{N} \Bigg],$$

where by convention we set $\hat{K}(-1/p) := \hat{K}(0)$. Note that with the above choice of ϵ_4 we have $\hat{K}\left(t_k^{(p)} - 1/N_1\right) \cdot \hat{K}(t_k^{(p)}) > 1 - 1/N$, provided that $\delta \in (0, \delta_0]$ and δ_0 is sufficiently small. We adopt in (2.28) a customary convention that the infimum of an empty set equals $+\infty$. The stopping time S_δ is triggered when the trajectory performs a sudden turn – this is undesirable as the trajectory may then return back to the region it has already visited and create correlations with the past.

For each $t \geq 0$, we denote by $\mathfrak{X}_t(\pi) := \bigcup_{0\leq s\leq t} X(s;\pi)$ the trace of the spatial component of the path π up to time t, and by $\mathfrak{X}_t(q;\pi) := [x : \text{dist}\,(x, \mathfrak{X}_t(\pi)) \leq 1/q]$ a tubular region around the path. We introduce the stopping time

$$U_\delta(\pi) := \inf\left[t \geq 0 : \exists\, k \geq 1 \text{ and } t \in [t_k^{(p)}, t_{k+1}^{(p)}) \text{ for which } X(t) \in \mathfrak{X}_{t_{k-1}^{(p)}}(q)\right]. \tag{2.29}$$

It is associated with the return of the X component of the trajectory to the tube around its past – this is again an undesirable way to create correlations with the past. Finally, we set the stopping time

$$\tau_\delta(\pi) := S_\delta(\pi) \wedge U_\delta(\pi). \tag{2.30}$$

2.3.1 The cut-off functions and the corresponding dynamics

Let $M > 0$ be fixed and p, q, N, N_1 be the positive integers defined in Section 2.3. We define now several auxiliary functions that will be used to introduce the cut-offs in the dynamics. These cut-offs will ensure that the particle moving under the modified dynamics will avoid self-intersections, will have no violent turns and the changes of its momentum will be under control. In addition, up to the stopping time τ_δ the motion of the particle will coincide with the motion under the original Hamiltonian flow.

Let $a_1 = 2$ and $a_2 = 3/2$. The functions $\psi_j : \mathbb{R}^d \times \mathbb{S}_1^{d-1} \to [0,1]$, $j = 1, 2$ are of C^∞ class and satisfy

$$\psi_j(\mathbf{k},\mathbf{l}) = \begin{cases} 1, & \text{if } \hat{\mathbf{k}}\cdot\mathbf{l} \geq 1 - 1/N \quad \text{and} \quad M_\delta^{-1} \leq |\mathbf{k}| \leq M_\delta \\ 0, & \text{if } \hat{\mathbf{k}}\cdot\mathbf{l} \leq 1 - a_j/N, \quad \text{or} \quad |\mathbf{k}| \leq (2M_\delta)^{-1}, \quad \text{or} \quad |\mathbf{k}| \geq 2M_\delta. \end{cases} \tag{2.31}$$

One can construct ψ_j in such a way that for arbitrary nonnegative integers m, n it is possible to find a constant $C_{m,n}$ for which $\|\psi_j\|_{m,n} \leq C_{m,n} N^{m+n}$. The cut-off function

$$\Psi(t,\mathbf{k};\pi) := \begin{cases} \psi_1\left(\mathbf{k}, \hat{K}\left(t_{k-1}^{(p)}\right)\right)\psi_2\left(\mathbf{k}, \hat{K}\left(t_k^{(p)} - 1/N_1\right)\right) & \text{for } t \in [t_k^{(p)}, t_{k+1}^{(p)}) \text{ and } k \geq 1 \\ \psi_2(\mathbf{k}, \hat{K}(0)) & \text{for } t \in [0, t_1^{(p)}) \end{cases} \tag{2.32}$$

will allow us to control the direction of the particle motion over each interval of the partition as well as not to allow the trajectory to escape to the regions where the change of the size of the velocity can be uncontrollable.

Let $\phi : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ be a function of the C^∞ class that satisfies $\phi(\mathbf{y},\mathbf{x}) = 1$, when $|\mathbf{y}-\mathbf{x}| \geq 3/q$ and $\phi(\mathbf{y},\mathbf{x}) = 0$, when $|\mathbf{y}-\mathbf{x}| \leq 2/q$. Again, in this case we can construct ϕ in such a way that $\|\phi\|_{m,n} \leq Cq^{m+n}$ for arbitrary integers m, n and a suitably chosen constant C. The function $\phi_k : \mathbb{R}^d \times \mathcal{C} \to [0,1]$ for a fixed path π is given by

$$\phi_k(\mathbf{y};\pi) = \prod_{0\leq l/q\leq t_{k-1}^{(p)}} \phi\left(\mathbf{y}, X\left(\frac{l}{q}\right)\right). \tag{2.33}$$

We set

$$\Phi(t,\mathbf{y};\pi) := \begin{cases} 1, & \text{if } 0 \le t < t_1^{(p)} \\ \phi_k(\mathbf{y};\pi), & \text{if } t_k^{(p)} \le t < t_{k+1}^{(p)}. \end{cases} \tag{2.34}$$

The function Φ shall be used to modify the dynamics of the particle in order to avoid a possibility of near self-intersections of its trajectory.

For a given $t \ge 0$, $(\mathbf{y},\mathbf{k}) \in \mathbb{R}_*^{2d}$ and $\pi \in \mathcal{C}$ let us denote $\Theta(t,\mathbf{y},\mathbf{k};\pi) := \Psi(t,\mathbf{k};\pi)\Phi\,(t,\mathbf{y};\pi)$. The following lemma can be verified by a direct calculation.

Lemma 2.3.1 *Let (β_1,β_2) be a multi-index with nonnegative integer valued components, $m = |\beta_1| + |\beta_2|$. There exists a constant C depending only on m and M such that*

$$|\partial_{\mathbf{y}}^{\beta_1}\partial_{\mathbf{k}}^{\beta_2}\Theta(t,\mathbf{y},\mathbf{k};\pi)| \le CT^{|\beta_1|}q^{2|\beta_1|}N^{|\beta_2|}$$

for all $t \in [0,T]$, $(\mathbf{y},\mathbf{k}) \in \mathcal{A}(2M)$, $\pi \in \mathcal{C}$.

Finally, let us set

$$F_\delta(t,\mathbf{y},\mathbf{l};\pi,\omega) = \Theta(t,\delta\mathbf{y},\mathbf{l};\pi)\nabla_{\mathbf{y}}H_1\,(\mathbf{y},|\mathbf{l}|;\omega)\,. \tag{2.35}$$

For a fixed $(\mathbf{x},\mathbf{k}) \in \mathbb{R}_*^{2d}$, $\delta > 0$ and $\omega \in \Omega$ we consider the modified particle dynamics with the cut-off that is described by the stochastic process $(y^{(\delta)}(t;\mathbf{x},\mathbf{k},\omega),\boldsymbol{l}^{(\delta)}(t;\mathbf{x},\mathbf{k},\omega))_{t\ge 0}$ whose paths are the solutions of the following equation

$$\begin{cases} \frac{dy^{(\delta)}(t;\mathbf{x},\mathbf{k})}{dt} = \left[H_0'(|\boldsymbol{l}^{(\delta)}(t;\mathbf{x},\mathbf{k})|) + \sqrt{\delta}\,\partial_l H_1\left(\frac{y^{(\delta)}(t;\mathbf{x},\mathbf{k})}{\delta},|\boldsymbol{l}^{(\delta)}(t;\mathbf{x},\mathbf{k})|\right)\right]\hat{\boldsymbol{l}}^{(\delta)}(t;\mathbf{x},\mathbf{k},) \\ \frac{d\boldsymbol{l}^{(\delta)}(t;\mathbf{x},\mathbf{k})}{dt} = -\frac{1}{\sqrt{\delta}}\,F_\delta\left(t,\frac{y^{(\delta)}(t;\mathbf{x},\mathbf{k})}{\delta},\boldsymbol{l}^{(\delta)}(t;\mathbf{x},\mathbf{k});y^{(\delta)}(\cdot;\mathbf{x},\mathbf{k}),\boldsymbol{l}^{(\delta)}(\cdot;\mathbf{x},\mathbf{k})\right) \\ \\ y^{(\delta)}(0;\mathbf{x},\mathbf{k}) = \mathbf{x}, \quad \boldsymbol{l}^{(\delta)}(0;\mathbf{x},\mathbf{k}) = \mathbf{k}. \end{cases} \tag{2.36}$$

We will denote by $\tilde{Q}_{\mathbf{x},\mathbf{k}}^{(\delta)}$ the law of the modified process $(y^{(\delta)}(\cdot;\mathbf{x},\mathbf{k}),\boldsymbol{l}^{(\delta)}(\cdot;\mathbf{x},\mathbf{k}))$ over $\mathcal{C}$ for a given $\delta > 0$ and by $\tilde{E}_{\mathbf{x},\mathbf{k}}^{(\delta)}$ the corresponding expectation. We assume that the initial momentum $\mathbf{k} \in A(M)$. From the construction of the cut-offs we immediately conclude that

$$\hat{\boldsymbol{l}}^{(\delta)}(t)\cdot\hat{\boldsymbol{l}}^{(\delta)}(t_{k-1}^{(p)}) \ge 1 - \frac{2}{N}, \quad t \in [t_{k-1}^{(p)},t_{k+1}^{(p)}), \quad \forall\, k \ge 0. \tag{2.37}$$

2.3.2 Some consequences of the mixing assumption

For any $t \ge 0$ we denote by $\mathcal{F}_t$ the σ-algebra generated by $(y^{(\delta)}(s),\boldsymbol{l}^{(\delta)}(s))$, $s \le t$. Here we suppress, for the sake of abbreviation, writing the initial data in the notation of the trajectory. In this section we assume that $M > 0$ is fixed, $X_1, X_2 : (\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^{d^2})^2 \to \mathbb{R}$ are certain continuous functions, Z is a random variable and g_1, g_2 are $\mathbb{R}^d\times[M^{-1},M]$-valued random vectors. We suppose further that Z, g_1, g_2, are $\mathcal{F}_t$-measurable, while $\tilde{X}_1, \tilde{X}_2$ are random fields of the form

$$\tilde{X}_i(\mathbf{x},k) = X_i\left(\left(\partial_k^j H_1(\mathbf{x},k),\nabla_{\mathbf{x}}\partial_k^j H_1(\mathbf{x},k),\nabla_{\mathbf{x}}^2\partial_k^j H_1(\mathbf{x},k)\right)_{j=0,1}\right).$$

For $i = 1,2$ we denote $g_i := (g_i^{(1)},g_i^{(2)})$ where $g_i^{(1)} \in \mathbb{R}^d$ and $g_i^{(2)} \in [M^{-1},M]$. We also let

$$U(\theta_1,\theta_2) := \mathbb{E}\left[\tilde{X}_1(\theta_1)\tilde{X}_2(\theta_2)\right], \quad \theta_1,\theta_2 \in \mathbb{R}^d\times[M^{-1},M]. \tag{2.38}$$

The following mixing lemma is useful in formalizing the "memory loss effect" and can be proved in the same way as Lemmas 5.2 and 5.3 of [6]. It is also similar in spirit to Lemma 1.3.2.

Lemma 2.3.2 *(i) Assume that $r, t \geq 0$ and*

$$\inf_{u \leq t} \left| g_i^{(1)} - \frac{y^{(\delta)}(u)}{\delta} \right| \geq \frac{r}{\delta}, \tag{2.39}$$

$\mathbb{P}$–a.s. on the set $Z \neq 0$ for $i = 1, 2$. Then, we have

$$\left| \mathbb{E}\left[\tilde{X}_1(g_1) \tilde{X}_2(g_2) Z \right] - \mathbb{E}\left[U(g_1, g_2) Z \right] \right| \leq 2\phi\left(\frac{r}{2\delta} \right) \|X_1\|_{L^\infty} \|X_2\|_{L^\infty} \|Z\|_{L^1(\Omega)}. \tag{2.40}$$

(ii) Let $\mathbb{E}X_1(\mathbf{0}, k) = 0$ for all $k \in [M^{-1}, M]$. Furthermore, we assume that g_2 satisfies (2.39),

$$\inf_{u \leq t} \left| g_1^{(1)} - \frac{y^{(\delta)}(u)}{\delta} \right| \geq \frac{r + r_1}{\delta} \tag{2.41}$$

and $|g_1^{(1)} - g_2^{(1)}| \geq r_1 \delta^{-1}$ for some $r_1 \geq 0$, $\mathbb{P}$-a.s. on the event $Z \neq 0$. Then, we have

$$\left| \mathbb{E}\left[\tilde{X}_1(g_1) \tilde{X}_2(g_2)\, Z \right] - \mathbb{E}\left[U(g_1, g_2) Z \right] \right| \leq C \phi^{1/2}\left(\frac{r}{2\delta} \right) \phi^{1/2}\left(\frac{r_1}{2\delta} \right) \|X_1\|_{L^\infty} \|X_2\|_{L^\infty} \|Z\|_{L^1(\Omega)} \tag{2.42}$$

for some absolute constant $C > 0$. Here the function U is given by (2.38).

2.3.3 The momentum diffusion

Let $\mathbf{k}(t)$ be a diffusion, starting at $k \in \mathbb{R}^d_*$ at $t = 0$, with the generator of the form

$$\begin{aligned} \mathcal{L}F(\mathbf{k}) &= \sum_{m,n=1}^{d} D_{mn}(\hat{\mathbf{k}}, |k|) \partial^2_{k_m, k_n} F(\mathbf{k}) + \sum_{m=1}^{d} E_m(\hat{\mathbf{k}}, |k|) \partial_{k_m} F(\mathbf{k}) \\ &= \sum_{m,n=1}^{d} \partial_{k_m} \left(D_{m,n}(\hat{\mathbf{k}}, |k|) \partial_{k_n} F(\mathbf{k}) \right), \quad F \in C_0^\infty(\mathbb{R}^d_*). \end{aligned} \tag{2.43}$$

Here the diffusion matrix is given by (2.12) and the drift vector is

$$E_m(\hat{\mathbf{k}}, l) = -\frac{1}{H_0'(l) l} \sum_{n=1}^{d} \int_0^{+\infty} s \frac{\partial^3 R(s\hat{\mathbf{k}}, l)}{\partial x_m \partial x_n^2}\, ds, \quad m = 1, \ldots, d.$$

Employing exactly the same argument as the one used in Section 4 of [6] it can be easily seen that this diffusion is supported on $\mathbb{S}^{d-1}_k$, where $k = |k|$. Moreover, it is non-degenerate on the sphere, for instance, under the assumption (2.9), cf. Proposition 4.3 of ibid.

Let $\mathfrak{Q}_{\mathbf{x},\mathbf{k}}$ be the law of the process $(\mathbf{x}(t), \mathbf{k}(t))$ that starts at $t = 0$ from $(\mathbf{x}, \mathbf{k})$ given by $\mathbf{x}(t) = x + \int_0^t H_0'(|\mathbf{k}(s)|) \hat{\mathbf{k}}(s) ds$, where $\mathbf{k}(t)$ is the diffusion described by (2.43). This process is a degenerate diffusion whose generator is given by

$$\tilde{\mathcal{L}}F(\mathbf{x}, \mathbf{k}) = \mathcal{L}_k F(\mathbf{x}, \mathbf{k}) + H_0'(|\mathbf{k}|)\, \hat{\mathbf{k}} \cdot \nabla_x F(\mathbf{x}, \mathbf{k}), \quad F \in C_0^\infty(\mathbb{R}^{2d}_*). \tag{2.44}$$

Here the notation $\mathcal{L}_k$ stresses that the operator $\mathcal{L}$ defined in (2.43) acts on the respective function in the k variable. We denote by $\mathfrak{M}_{x,k}$ the expectation corresponding to the path measure $\mathfrak{Q}_{\mathbf{x},\mathbf{k}}$.

2.3.4 The augmented process

The following construction of the augmentation of path measures has been carried out in Section 6.1 of [58]. Let $s \ge 0$ be fixed and $\pi \in \mathcal{C}$. Then, according to Lemma 6.1.1 of ibid. there exists a unique probability measure, that is denoted by $\delta_\pi \otimes_s \mathfrak{Q}_{X(s),K(s)}$, such that for any pair of events $A \in \mathcal{M}^s$, $B \in \mathcal{M}$ we have $\delta_\pi \otimes_s \mathfrak{Q}_{X(s),K(s)}[A] = \mathbf{1}_A(\pi)$ and $\delta_\pi \otimes_s \mathfrak{Q}_{X(s),K(s)}[\theta_s(B)] = \mathfrak{Q}_{X(s),K(s)}[B]$. The following result is a direct consequence of Theorem 6.2.1 of [58].

Proposition 2.3.3 *There exists a unique probability measure $R^{(\delta)}_{x,k}$ on $\mathcal{C}$ such that $R^{(\delta)}_{x,k}[A] := Q^{(\delta)}_{x,k}[A]$ for all $A \in \mathcal{M}^{\tau_\delta}$ and the regular conditional probability distribution of $R^{(\delta)}_{x,k}[\,\cdot\,|\mathcal{M}^{\tau_\delta}]$ is given by $\delta_\pi \otimes_{\tau_\delta(\pi)} \mathfrak{Q}_{X(\tau_\delta(\pi)),K(\tau_\delta(\pi))}$, $\pi \in \mathcal{C}$. This measure shall be also denoted by $Q^{(\delta)}_{x,k} \otimes_{\tau_\delta} \mathfrak{Q}_{X(\tau_\delta),K(\tau_\delta)}$.*

Note that for any $(\mathbf{x},\mathbf{k}) \in \mathcal{A}(M)$ and $A \in \mathcal{M}^{\tau_\delta}$ we have

$$R^{(\delta)}_{\mathbf{x},\mathbf{k}}[A] = Q^{(\delta)}_{\mathbf{x},\mathbf{k}}[A] = \tilde{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}[A], \tag{2.45}$$

that is, the law of the augmented process coincides with that of the true process, and of the modified process with the cut-offs until the stopping time τ_δ. Hence, according to the uniqueness part of Proposition 2.3.3, in such a case $Q^{(\delta)}_{x,k} \otimes_{\tau_\delta} \mathfrak{Q}_{X(\tau_\delta),K(\tau_\delta)} = \tilde{Q}^{(\delta)}_{x,k} \otimes_{\tau_\delta} \mathfrak{Q}_{X(\tau_\delta),K(\tau_\delta)}$. We denote by $E^{(\delta)}_{x,k}$ the expectation with respect to the augmented measure described by the above proposition. Let also $R^{(\delta)}_{x,k,\pi}$, $E^{(\delta)}_{x,k,\pi}$ denote the respective conditional law and expectation obtained by conditioning $R^{(\delta)}_{x,k}$ on $\mathcal{M}^{\tau_\delta}$.

The following proposition is of crucial importance for us, as it shows that the law of the augmented process is close to that of the momentum diffusion as $\delta \to 0$. To abbreviate the notation we let

$$N_t(G) := G(t, X(t), K(t)) - G(0, X(0), K(0)) - \int_0^t (\partial_\varrho + \tilde{\mathcal{L}}) G(\varrho, X(\varrho), K(\varrho)))\, d\varrho$$

for any $G \in C^{1,1,3}_b([0,+\infty) \times \mathbb{R}^{2d}_*)$ and $t \ge 0$.

Proposition 2.3.4 *Suppose that $(x,k) \in \mathcal{A}(M)$ and $\zeta \in C_b((\mathbb{R}^{2d}_*)^n)$ is nonnegative. Let $\gamma_0 \in (0,1/2)$ and let $0 \le t_1 < \cdots < t_n \le T_* \le t < v \le T$. We assume further that $v - t \ge \delta^{\gamma_0}$. Then, there exist constants γ_1, C such that for any function $G \in C^{1,1,3}([T_*,T] \times \mathbb{R}^{2d}_*)$ we have*

$$\left| E^{(\delta)}_{x,k} \left\{ [N_v(G) - N_t(G)]\, \tilde{\zeta} \right\} \right| \le C \delta^{\gamma_1} (v - t) \|G\|^{[T_*,T]}_{1,1,3} T^2 E^{(\delta)}_{x,k} \tilde{\zeta}. \tag{2.46}$$

Here $\tilde{\zeta}(\pi) := \zeta(X(t_1), K(t_1), \ldots, X(t_n), K(t_n))$, $\pi \in \mathcal{C}(T,\delta)$. The choice of the constants γ_1, C does not depend on (x,k), $\delta \in (0,1]$, ζ, times $t_1, \ldots, t_n, T_*, T, v, t$, or the function G.

Proof. Let $0 = s_0 \le s_1 \le \ldots \le s_n \le t$ and $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R}^{2d}_*)$ be Borel sets. We denote $A_0 := \mathcal{C}$ and for any $k \in \{1, \ldots, n\}$, $s \le s_k$ we define the events

$$A_k := [\pi :\ (X(s_1), K(s_1)) \in B_1, \ldots, (X(s_k), K(s_k)) \in B_k]$$

and their shifted counterparts

$$A^{(s)}_k := [\pi :\ (X(s_k - s), K(s_k - s)) \in B_k, \ldots, (X(s_n - s), K(s_n - s)) \in B_n].$$

For $(\mathbf{x},\mathbf{k})\in\mathbb{R}^{2d}_*$, $\pi\in\mathcal{C}$ and $G\in C^{1,1,2}([0,+\infty)\times\mathbb{R}^{2d}_*)$ we let

$$\begin{aligned}\widehat{\mathcal{L}}_t G(t,x,\mathbf{k};\pi) &:= H_0'(|k|)\,\hat{\mathbf{k}}\cdot\nabla_x G(t,\mathbf{x},\mathbf{k}) + \Theta^2(t,X(t),K(t);\pi)\mathcal{L}_k G(t,x,\mathbf{k})\\ &\quad - \Theta(t,X(t),K(t);\pi)\sum_{m,n=1}^{d}\partial_{K_m}\Theta(t,X(t),K(t);\pi)D_{m,n}(\hat{\mathbf{k}},|k|)\partial_{k_n}G(t,x,\mathbf{k})\end{aligned}$$

and

$$\widehat{N}_t(G) := G(t,X(t),K(t)) - G(0,X(0),K(0)) - \int_0^t (\partial_\varrho + \widehat{\mathcal{L}}_\varrho)G(\varrho,X(\varrho),K(\varrho);\pi)\,d\varrho.$$

It follows from the definition of the stopping time $\tau_\delta(\pi)$ and the cut-off function Θ that

$$\nabla_K\Theta(t,X(t),K(t);\pi) = \mathbf{0},\quad t\in[0,\tau_\delta(\pi)],$$

hence

$$\widehat{\mathcal{L}}_t G(t,X(t),K(t);\pi) = \tilde{\mathcal{L}}G(t,X(t),K(t);\pi),\quad t\in[0,\tau_\delta(\pi)].$$

We need the following result.

Lemma 2.3.5 *Suppose that $(x,k)\in\mathcal{A}(M)$ and $\zeta\in C_b((\mathbb{R}^{2d}_*)^n)$ is nonnegative. Let $\gamma_0'\in(0,1)$, $0\le t_1<\cdots<t_n\le T_*\le t<v\le T$ and $t-T_*\ge\delta^{\gamma_0'}$. Then, there exist constants γ_1', $C'>0$ such that for any function $G\in C^{1,1,3}([T_*,T]\times\mathbb{R}^{2d}_*)$ we have*

$$\left|\tilde{E}^{(\delta)}_{x,k}\left\{[\widehat{N}_v(G)-\widehat{N}_t(G)]\tilde{\zeta}\right\}\right| \le C'\delta^{\gamma_1'}(v-t)\|G\|^{[T_*,T]}_{1,1,3}T^2\tilde{E}^{(\delta)}_{x,k}\tilde{\zeta}. \tag{2.47}$$

The choice of the constants γ_1', C' does not depend on (x,k), $\delta\in(0,1]$, times $t_1,\ldots,t_n,T_*,T,v,t$, or function G.

The proof of this lemma follows very closely the argument presented in Section 5.3 of [6] and we postpone it until Section 2.6. In the meantime we apply this result to conclude the proof of Proposition 2.3.4. We write

$$\begin{aligned}&E^{(\delta)}_{x,\mathbf{k},\pi}[N_v(G)-N_{v\wedge\tau_\delta(\pi)}(G),A_n]\\ &\quad=\sum_{p=0}^{n-1}\mathbf{1}_{[s_p,s_{p+1})}(\tau_\delta(\pi))\mathbf{1}_{A_p}(\pi)\mathfrak{M}_{X(\tau_\delta(\pi)),K(\tau_\delta(\pi))}[N_{v-\tau_\delta(\pi)}(G),A^{(\tau_\delta(\pi))}_{p+1}]\\ &\quad+\mathbf{1}_{[s_n,v)}(\tau_\delta(\pi))\mathbf{1}_{A_n}(\pi)\mathfrak{M}_{X(\tau_\delta(\pi)),K(\tau_\delta(\pi))}[N_{v-\tau_\delta(\pi)}(G)].\end{aligned} \tag{2.48}$$

When $\tau_\delta(\pi)\in[s_p,s_{p+1})$ we obviously have

$$\mathfrak{M}_{X(\tau_\delta(\pi)),K(\tau_\delta(\pi))}[N_{v-\tau_\delta(\pi)}(G),A^{(\tau_\delta(\pi))}_{p+1}] = \mathfrak{M}_{X(\tau_\delta(\pi)),K(\tau_\delta(\pi))}[N_{t-\tau_\delta(\pi)}(G),A^{(\tau_\delta(\pi))}_{p+1}]$$

and $\mathfrak{M}_{X(\tau_\delta(\pi)),K(\tau_\delta(\pi))}[N_{v-\tau_\delta(\pi)}(G)]=0$. Hence the left hand side of (2.48) equals

$$\sum_{p=0}^{n-1}\mathbf{1}_{[s_p,s_{p+1})}(\tau_\delta(\pi))\mathbf{1}_{A_p}(\pi)\mathfrak{M}_{X(\tau_\delta(\pi)),K(\tau_\delta(\pi))}[N_{t-\tau_\delta(\pi)}(G),A^{(\tau_\delta(\pi))}_{p+1}] \tag{2.49}$$

$$= E^{(\delta)}_{x,\mathbf{k},\pi}[N_t(G)-N_{t\wedge\tau_\delta(\pi)}(G),A_n].$$

We conclude from (2.48), (2.49) that

$$E^{(\delta)}_{x,\mathbf{k},\pi}[N_v(G), A_n] = E^{(\delta)}_{x,\mathbf{k},\pi}[N_{v\wedge\tau_\delta(\pi)}(G) + N_t(G) - N_{t\wedge\tau_\delta(\pi)}(G), A_n] \qquad (2.50)$$
$$= E^{(\delta)}_{x,\mathbf{k},\pi}[N_{(v\wedge\tau_\delta(\pi))\vee t}(G), A_n]$$

and therefore

$$E^{(\delta)}_{x,\mathbf{k}}[N_v(G), A_n] = E^{(\delta)}_{x,\mathbf{k}}\left[E^{(\delta)}_{x,\mathbf{k},\pi}[N_{(v\wedge\tau_\delta(\pi))\vee t}(G), A_n]\right] \qquad (2.51)$$
$$= E^{(\delta)}_{x,\mathbf{k}}\left[E^{(\delta)}_{x,\mathbf{k},\pi}\left[N_{(v\wedge\tau_\delta(\pi))\vee t}(G), A_n\right], \tau_\delta(\pi) \le t\right]$$
$$+ E^{(\delta)}_{x,\mathbf{k}}\left[E^{(\delta)}_{x,\mathbf{k},\pi}\left[N_{(v\wedge\tau_\delta(\pi))\vee t}(G), A_n\right], \tau_\delta(\pi) > t\right].$$

The first term on the utmost right hand side of (2.51) equals $E^{(\delta)}_{x,\mathbf{k}}[N_t(G), A_n, \tau_\delta \le t]$, while the second one equals $\tilde{E}^{(\delta)}_{x,\mathbf{k}}\left[N_{(v\wedge\tau_\delta)\vee t}(G), B\right]$. Here $B := A_n \cap [\tau_\delta > t]$ is an $\mathcal{M}_t$–measurable event. Suppose that $\gamma_0' \in (\gamma_0 + 1/2, 1)$ and let $L := [\delta^{-\gamma_0'}]$ be yet another mesh size parameter. We define

$$\sigma := L^{-1}[([L(v\wedge\tau_\delta)] + 2) \vee ([Lt] + 2)]$$

and note that

$$\tilde{E}^{(\delta)}_{x,\mathbf{k}}[N_\sigma(G), B] = \sum_{p=[Lt]+2}^{[Lv]+2} \tilde{E}^{(\delta)}_{x,\mathbf{k}}\left[N_{p/L}(G), B, \sigma = \frac{p}{L}\right] \qquad (2.52)$$

Representing the event $[\sigma = p/L]$ as the difference of $[\sigma \ge p/L]$ and $[\sigma \ge (p+1)/L]$ (note that $[\sigma \ge ([Lv]+3)/L] = \emptyset$) and grouping the terms of the sum that correspond to the same index p we obtain that the right hand side of (2.52) equals

$$\tilde{E}^{(\delta)}_{x,\mathbf{k}}\left[N_{([Lt]+2)/L}(G), B\right] + \sum_{p=[Lt]+2}^{[Lv]+2} \tilde{E}^{(\delta)}_{x,\mathbf{k}}\left[N_{p+1/L}(G) - N_{p/L}(G), B, \sigma \ge \frac{p+1}{L}\right]. \qquad (2.53)$$

Since the event $B \cap [\sigma \ge (p+1)/L]$ is $\mathcal{M}^{(p-1)/L}$-measurable, from Lemma 2.3.5 we conclude that the absolute value of each term appearing under the summation sign in (2.53) can be estimated by $C'\|G\|_{1,1,3}\delta^{\gamma_1'}L^{-1}\tilde{Q}^{(\delta)}_{x,\mathbf{k}}[B]$ which implies

$$\left|\tilde{E}^{(\delta)}_{x,\mathbf{k}}[N_\sigma(G), B] - \tilde{E}^{(\delta)}_{x,\mathbf{k}}\left[N_{([Lt]+2)L^{-1}}(G), B\right]\right| \le C'\delta^{\gamma_1'}\|G\|^{[T_*,T]}_{1,1,3}T^2\,\tilde{Q}^{(\delta)}_{x,\mathbf{k}}[B]\,\frac{[Lv]+1-[Lt]}{L}.$$

A direct calculation using formulas (2.26) allows us to conclude also that both $|N_\sigma(G) - N_{(v\wedge\tau_\delta)\vee t}(G)|$ and $|N_{([Lt]+2)L^{-1}}(G) - N_t(G)|$ are estimated by $C\|G\|^{[T_*,T]}_{1,1,3}\delta^{\gamma_0'-1/2}$. Hence, (since $\gamma_0' > 1/2 + \gamma_0$)

$$\left|\tilde{E}^{(\delta)}_{x,\mathbf{k}}\left[N_{(v\wedge\tau_\delta)\vee t}(G), B\right] - \tilde{E}^{(\delta)}_{x,\mathbf{k}}[N_t(G), B]\right| \le \left|\tilde{E}^{(\delta)}_{x,\mathbf{k}}\left[N_\sigma(G) - N_{(v\wedge\tau_\delta)\vee t}(G), B\right]\right|$$
$$+ \left|\tilde{E}^{(\delta)}_{x,\mathbf{k}}[N_\sigma(G), B] - \tilde{E}^{(\delta)}_{x,\mathbf{k}}\left[N_{([Lt]+2)L^{-1}}(G), B\right]\right| + \left|\tilde{E}^{(\delta)}_{x,\mathbf{k}}\left[N_{([Lt]+2)L^{-1}}(G) - N_t, B\right]\right|$$
$$\le C\delta^{\gamma_1}\|G\|^{[T_*,T]}_{1,1,3}T^2\tilde{Q}^{(\delta)}_{x,\mathbf{k}}[B]\,(v-t)\vee\delta^{\gamma_0} \qquad (2.54)$$

for a certain constant $C > 0$ and $\gamma_1 := \min[\gamma_0' - \gamma_0 - 1/2, \gamma_1']$. From (2.51), (2.54) and the observation just below (2.51), we obtain

$$\left|E^{(\delta)}_{x,\mathbf{k}}[N_v(G) - N_t(G), A_n]\right| \le C\delta^{\gamma_1}\|G\|^{[T_*,T]}_{1,1,3}T^2R^{(\delta)}_{x,\mathbf{k}}[A_n]\,(v-t)\vee\delta^{\gamma_0}$$

for a certain constant $C > 0$ and the conclusion of Proposition 2.3.4 follows. □

2.3.5 An estimate of the stopping time

The purpose of this section is to prove the following estimate for $R^{(\delta)}_{\mathbf{x},\mathbf{k}}[\tau_\delta < T]$.

Theorem 2.3.6 *Assume that the dimension $d \ge 3$. Then, one can choose $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$ in such a way that there exist constants $C, \gamma > 0$ for which*

$$R^{(\delta)}_{x,k}[\tau_\delta < T] \le C\delta^\gamma T, \quad \forall\, \delta \in (0,1],\ T \ge 1,\ (x,k) \in \mathcal{A}(M). \tag{2.55}$$

Proof. We obviously have

$$[\tau_\delta < T] = [U_\delta \le \tau_\delta,\ U_\delta < T] \cup [S_\delta \le \tau_\delta,\ S_\delta < T] \tag{2.56}$$

with the stopping times S_δ and U_δ defined in (2.28) and (2.29). Let us denote the first and second event appearing on the right hand side of (2.56) by $A(\delta)$ and $B(\delta)$ respectively. To show that (2.56) holds we prove that the $R^{(\delta)}_{x,k}$ probabilities of both events can be estimated by $C\delta^\gamma T$ for some $C, \gamma > 0$: see (2.64), (2.65) and (2.69).

An estimate of $R^{(\delta)}_{x,k}[A(\delta)]$

The first step towards obtaining the desired estimate will be to replace the event $A(\delta)$ whose definition involves a stopping time by an event $C(\delta)$ whose definition depends only on deterministic times, see (2.57) below. Next we use the estimate (2.46) of Proposition 2.3.4 for an appropriately chosen function G to reduce the question of bounding the $R^{(\delta)}_{x,k}$ probability of $\tilde{A}(\delta)$ by an easier problem of estimating its $\mathfrak{Q}_{x,k}$ probability ($\mathfrak{Q}_{x,k}$ corresponds to a degenerate diffusion determined by (2.44)). The latter is achieved by using bounds on heat kernels corresponding to hypoelliptic diffusions due to Kusuoka and Stroock.

We assume in this section to simplify the notation and without any loss of generality that $h^*(M) = 1$. Note that then

$$A(\delta) \subset \tilde{A}(\delta) := \left[\left|X\left(\frac{j}{q}\right) - X\left(\frac{i}{q}\right)\right| \le \frac{3}{q} :\ 1 \le i \le j \le [Tq], \quad |i-j| \ge \frac{q}{p}\right] \tag{2.57}$$

and thus

$$R^{(\delta)}_{x,k}[A(\delta)] \le [Tq]^2 \max\left\{R^{(\delta)}_{x,k}\left[\left|X\left(\frac{j}{q}\right) - X\left(\frac{i}{q}\right)\right| \le \frac{3}{q}\right] :\ 1 \le i \le j \le [Tq],\ |i-j| \ge \frac{q}{p}\right\}. \tag{2.58}$$

Suppose that $f^{(\delta)} : \mathbb{R}^d \to [0,1]$ is a C^∞–regular function that satisfies $f(x) = 1$, if $|x| \le 4h/q$ and $f^{(\delta)}(x) = 0$, if $|x| \ge 5/q$. We assume furthermore that i, j are positive integers such that $(j-i)/q \in [0,1]$ and $\|f^{(\delta)}\|_3 \le 2q^3$. For any $x_0 \in \mathbb{R}^d$ and $i/q \le t \le j/q$ define

$$G_j(t,x,k;x_0) := \mathfrak{M}_{x,k} f^{(\delta)}\left(X\left(\frac{j}{q} - t\right) - x_0\right).$$

Obviously, we have

$$\partial_t G_j(t,x,k;x_0) + \tilde{\mathcal{L}} G_j(t,x,k;x_0) = 0.$$

Hence, using Proposition 2.3.4 with $v = j/q$ and $t = i/q$ (note that $v - t \ge 1/p \ge \delta^{\epsilon_2}$ and $\epsilon_2 \in (0,1/2)$), we obtain that there exists $\gamma_1 > 0$ such that

$$\left|E^{(\delta)}_{x,k}\left[f^{(\delta)}\left(X\left(\frac{j}{q}\right) - x_0\right) - G_j\left(\frac{i}{q}, X\left(\frac{i}{q}\right), K\left(\frac{i}{q}\right); x_0\right) \,\Big|\, \mathcal{M}^{i/q}\right]\right| \tag{2.59}$$
$$\le C\,\frac{j-i}{q}\,\|G_j(\cdot,\cdot,\cdot;x_0)\|^{[i/q,j/q]}_{1,1,3} T^2\delta^{\gamma_1}, \quad \forall\, \delta \in (0,1].$$

According to [57] Theorem 2.58, p. 53 we have

$$\|G_j(\cdot,\cdot,\cdot;x_0)\|_{1,1,3}^{[i/q,j/q]} \le C\|f^{(\delta)}\|_3 \le Cq^3 \le C\delta^{-3(\epsilon_2+\epsilon_3)}, \quad j\in\{0,\ldots,[qT]\}. \tag{2.60}$$

Hence combining (2.59) and (2.60) we obtain that the left hand side of (2.59) is less than, or equal to $C\,\delta^{\gamma_1-3(\epsilon_2+\epsilon_3)}$ for all $\delta\in(0,1]$. Let now $i_0 = j-\dfrac{q}{p}$ so that $1\le i\le i_0\le j\le[Tq]$. We have

$$\begin{aligned} R_{x,k}^{(\delta)}\left[\left|X\left(\frac{j}{q}\right)-X\left(\frac{i}{q}\right)\right|\le\frac{3}{q}\right] &\le E_{x,k}^{(\delta)}\left[f^{(\delta)}\left(X\left(\frac{j}{q}\right)-X\left(\frac{i}{q}\right)\right)\right] \\ &= E_{x,k}^{(\delta)}\left[E_{x,k}^{(\delta)}\left[f^{(\delta)}\left(X\left(\frac{j}{q}\right)-\mathbf{y}\right)\,\Big|\,\mathcal{M}^{i_0/q}\right]\Big|_{\mathbf{y}=X(i/q)}\right]. \end{aligned} \tag{2.61}$$

According to (2.59) and (2.60) we can estimate the utmost right hand side of (2.61) by

$$\sup_{x,\mathbf{y},k}\left\{\mathfrak{M}_{x,k}f^{(\delta)}\left(X\left(\frac{1}{p}\right)-\mathbf{y}\right): x,\mathbf{y}\in\mathbb{R}^d,\ k\in A(2M)\right\}+C\,\delta^{\gamma_1-3(\epsilon_2+\epsilon_3)}T^2. \tag{2.62}$$

To estimate the first term in (2.62) we use the following.

Lemma 2.3.7 *Let p,q be as in (2.27). Then, there exist positive constants C_1, C_2 and C_3 such that for all $x,\mathbf{y}\in\mathbb{R}^d$, $k\in A(2M)$, $j\in\{1,\ldots,[pT]\}$, $\delta\in(0,1]$ we have*

$$\mathfrak{Q}_{x,k}\left[\left|X\left(\frac{j}{p}\right)-\mathbf{y}\right|\le\frac{5}{q}\right]\le C_1\left(\frac{p^{C_2}}{q^d}+e^{-C_3p}\right). \tag{2.63}$$

We postpone the proof of the lemma for a moment in order to finish the estimate of $R_{x,k}^{(\delta)}[A(\delta)]$. Using (2.63) we obtain that the expression in (2.62) can be estimated by

$$C_1\left(\frac{p^{C_2}}{q^d}+e^{-C_3p}\right)+C\,\delta^{\gamma_1-3(\epsilon_2+\epsilon_3)}T^2 \le C_1\delta^{(d-C_2)\epsilon_2+d\epsilon_3}+\exp\left\{-C_3\delta^{-\epsilon_2}\right\}+C\,\delta^{\gamma_1-3(\epsilon_2+\epsilon_3)}T^2.$$

Hence, from (2.58), we obtain that

$$\begin{aligned} &R_{x,k}^{(\delta)}[A(\delta)] \le [Tq]^2\left(C_1\delta^{(d-C_2)\epsilon_2+d\epsilon_3}+\exp\left\{-C_3\delta^{-\epsilon_2}\right\}+C\,\delta^{\gamma_1-3(\epsilon_2+\epsilon_3)}T^2\right) \\ &\le CT^2\left(\delta^{(d-2-C_1)\epsilon_2+(d-2)\epsilon_3}+\delta^{-2(\epsilon_2+\epsilon_3)}\exp\left\{-C_3\delta^{-\epsilon_2}\right\}+\delta^{\gamma_1-5(\epsilon_2+\epsilon_3)}T^2\right)\le C\delta^{\gamma_2}T^4 \end{aligned} \tag{2.64}$$

for $\gamma_2 := \min[(d-2-C_1)\epsilon_2+(d-2)\epsilon_3, \gamma_1-5(\epsilon_2+\epsilon_3)]>0$, provided that $\epsilon_2+\epsilon_3<\gamma_1/5$ and $\epsilon_2\in(0,(d-2)\epsilon_3/(C_1+2-d))$. Here with no loss of generality we have assumed that $C_1+2>d$. Recall also that $d\ge3$. Now suppose that $\gamma_3\in(0,\gamma_2)$. Consider two cases: $T^3<\delta^{-\gamma_3}$ and $T^3\ge\delta^{-\gamma_3}$. In the first one, the utmost right hand side of (2.64) can be bound from above by $C\delta^{\gamma_2-\gamma_3}T$. In the second we have a trivial bound of the left side by $\delta^{\gamma_3/3}T$. We have proved therefore that

$$R_{x,k}^{(\delta)}[A(\delta)]\le C\delta^{\gamma}T \tag{2.65}$$

for some $C,\gamma>0$ independent of δ and T.

The proof of Lemma 2.3.7**.** We prove this lemma by induction on j. First, we verify it for $j=1$. Without any loss of generality we may suppose that $k=(k_1,\ldots,k_d)$ and

$k_d > (4dM_\delta)^{-1}$. Let $\tilde{D}_{mn} : \mathbb{R}^{d-1} \to \mathbb{R}$, $m, n = 1, \ldots, d-1$, $\tilde{E}_m : \mathbb{R}^{d-1} \to \mathbb{R}$, $m = 1, \ldots, d$ be given by

$$\tilde{D}_{pq}(\mathbf{l}) := D_{pq}(k^{-1}\mathbf{l}, k^{-1}\sqrt{k^2 - l^2}, k), \quad \tilde{E}_p(\mathbf{l}) := E_p(k^{-1}\mathbf{l}, k^{-1}\sqrt{k^2 - l^2}, k),$$

when $\mathbf{l} \in Z := [\mathbf{l} \in \mathbb{B}_k^{d-1} : k^{-1}\sqrt{k^2 - l^2} > (4dM_\delta)^{-1}]$, $l = |\mathbf{l}|$. These functions are C^∞ smooth and bounded together with all their derivatives. Note also that the matrix $\tilde{\mathbf{D}} = [\tilde{D}_{mn}]$ is symmetric and $\tilde{\mathbf{D}}\xi \cdot \xi \geq \lambda_0 |\xi|^2$ for all $\xi \in \mathbb{R}^{d-1}$ and a certain $\lambda_0 > 0$. The projection $K(t) = (K_1(t), \ldots, K_d(t))$ of the canonical path process $(X(t;\pi), K(t;\pi))$ considered over the probability space $(\mathcal{C}, \mathcal{M}, \mathfrak{Q}_{x,k_0})$, where $k_0 := (\mathbf{l}, \sqrt{k^2 - l^2})$, with $\mathbf{l} \in Z$, is a diffusion whose generator equals $\mathcal{L}$, see (2.43). It can be easily seen that $(K_1(t), \ldots, K_{d-1}(t))_{t \geq 0}$, is then a diffusion starting at $\mathbf{l}$, whose generator $\mathcal{N}$ is of the form

$$\mathcal{N}F(\mathbf{l}, \mathbf{x}) := \sum_{p=1}^{d-1} X_p^2 F(\mathbf{l}) + \sum_{q=1}^{d-1} a_q(\mathbf{l}) \partial_{l_q} F(\mathbf{l}), \quad F \in C_0^\infty(\mathbb{R}^{d-1}), \tag{2.66}$$

where $a_q(\mathbf{l})$, $q = 1, \ldots, d-1$ are certain C^∞-functions and

$$X_p(\mathbf{l}) := \sum_{q=1}^{d-1} \tilde{D}_{pq}^{1/2}(\mathbf{l}) \partial_{l_q}, \quad p = 1, \ldots, d-1.$$

The $(d-1) \times (d-1)$ matrix $[\tilde{D}_{pq}^{1/2}(\mathbf{l})]$ is non-degenerate when $\mathbf{l} \in Z$. Let

$$\tilde{\mathcal{N}}F(\mathbf{l}, \mathbf{x}) := \sum_{p=1}^{d-1} \tilde{X}_p^2 F(\mathbf{l}, \mathbf{x}) + \tilde{X}_0 F(\mathbf{l}, \mathbf{x}), \quad F \in C_0^\infty(\mathbb{R}^{d-1} \times \mathbb{R}^d),$$

where $\tilde{X}_0$ is a C^∞–smooth extension of the field

$$X_0(\mathbf{l}) := \frac{H_0'(k)}{k} \sum_{q=1}^{d-1} l_q \partial_{x_q} + \frac{H_0'(k)}{k} \sqrt{k^2 - l^2}\, \partial_{x_d} + \sum_{q=1}^{d-1} a_q(\mathbf{l}) \partial_{l_q}, \quad \mathbf{l} \in Z.$$

It can be shown, by the same type of argument as that given on pp. 122-123 of [6], that for each $(x, \mathbf{l})$, with $\mathbf{l} \in Z$, the linear space spanned at that point by the fields belonging to the Lie algebra generated by $X_0, \ldots, X_{d-1}$ is of dimension $2d-1$. One can also guarantee that the extensions $\tilde{X}_0, \ldots, \tilde{X}_{d-1}$ satisfy the same condition. We shall denote the respective extension of $\mathcal{N}$ by the same symbol.

Set $\mathbf{l}_0 := (k_1, \ldots, k_{d-1})$. Let $\mathcal{R}_{\mathbf{l}_0}$, $\tilde{\mathcal{R}}_{\mathbf{x},\mathbf{l}_0}$ be the path measures supported on $\mathcal{C}^{d-1}$ and $\mathcal{C}^{d,d-1}$ respectively that solve the martingale problems corresponding to the generators $\mathcal{N}$ and $\tilde{\mathcal{N}}$ with the respective initial conditions at $t = 0$ given by $\mathbf{l}_0$ and $(\mathbf{x}, \mathbf{l}_0)$. Let $r(t, x - \mathbf{y}, \mathbf{l}_1, \mathbf{l}_2)$, $t \in (0, +\infty)$, $x, \mathbf{y} \in \mathbb{R}^d$, $\mathbf{l}_1, \mathbf{l}_2 \in \mathbb{R}^{d-1}$ be the transition of probability density that corresponds to $\tilde{\mathcal{R}}_{\mathbf{x},\mathbf{l}_0}$. Using Corollary 3.25 p. 22 of [47] we have that for some constants $C, m > 0$

$$r\,(t, \mathbf{y}, k, \mathbf{l}) \leq Ct^{-m}, \quad \forall\, \mathbf{y} \in \mathbb{R}^d,\, k,\, \mathbf{l} \in \mathbb{R}^{d-1},\, t \in (0, 1]. \tag{2.67}$$

Denote by $\tau_Z(\pi)$ the exit time of a path $\pi \in \mathcal{C}^{d-1}$ from the set Z. For any $\pi \in \mathcal{C}^{d,d-1}$ we set also $\tilde{\tau}_Z(\pi) = \tau_Z(K(\cdot;\pi))$. Let $S : \mathbb{B}_k^{d-1} \to \mathbb{S}_k^{d-1}$ be given by

$$S(\mathbf{l}) := (l_1, \ldots, l_{d-1}, \sqrt{k^2 - l^2}), \quad \mathbf{l} = (l_1, \ldots, l_{d-1}) \in \mathbb{B}_k^{d-1},\, l := |\mathbf{l}|$$

and let $\tilde{S} : \mathcal{C}^{d,d-1} \to \mathcal{C}$ be given by $\tilde{S}(\pi)(t) := (X(t;\pi), S \circ K(t;\pi))$, $t \geq 0$. For any $A \in \mathcal{M}^{\tilde{\tau}_Z}$ we have $\tilde{\mathcal{R}}_{\mathbf{x},\mathbf{l}_0}[\tilde{S}^{-1}(A)] = \mathfrak{Q}_{\mathbf{x},S(\mathbf{l}_0)}[A]$. Since the event $[|X(1/p) - \mathbf{y}| \leq 5/q] \cap [\tilde{\tau}_Z \geq 1/p]$ is $\mathcal{M}^{\tilde{\tau}_Z}$–measurable we have

$$\mathfrak{Q}_{x,k}\left[\left|X\left(\frac{1}{p}\right) - \mathbf{y}\right| \leq \frac{5}{q}\right] \leq \tilde{\mathcal{R}}_{x,\mathbf{l}_0}\left[\left|X\left(\frac{1}{p}\right) - \mathbf{y}\right| \leq \frac{5}{q}, \tilde{\tau}_Z \geq \frac{1}{p}\right] + \mathcal{R}_{\mathbf{l}_0}\left[\tau_Z < \frac{1}{p}\right] \tag{2.68}$$

$$\leq C\bar{\omega}_d p^m \left(\frac{4}{q}\right)^d + Ce^{-C_3 p}.$$

Here $\bar{\omega}_d$ denotes the volume of $\mathbb{B}^d$. To obtain the last inequality we have used (2.67) and an estimate for non-degenerate diffusions stating that $\mathcal{R}_{\mathbf{l}_0}[\tau_Z < 1/p] < Ce^{-C_3 p}$ for some constants $C, C_3 > 0$ depending only on d and λ_0, see e.g. (2.1) p. 87 of [58]. Inequality (2.68) implies easily (2.63) for $j = 1$ with $C_1 = m$. To finish the induction argument assume that (2.63) holds for a certain j. We show that it holds for $j + 1$ with the same constants C_1, C_2 and $C_3 > 0$. The latter follows easily from the Chapman-Kolmogorov equation, since

$$\mathfrak{Q}_{x,k}\left[\left|X\left(\frac{j+1}{p}\right) - \mathbf{y}\right| \leq \frac{5}{q}\right] = \iint\limits_{\mathbb{R}^d \times \mathbb{S}_k^{d-1}} \mathfrak{Q}_{\mathbf{y},\mathbf{l}}\left[\left|X\left(\frac{j}{p}\right) - \mathbf{y}\right| \leq \frac{5}{q}\right] Q\left(\frac{1}{p}, x, k, d\mathbf{y}, d\mathbf{l}\right)$$

$$\stackrel{\text{induction assumpt.}}{\leq} C_1\left[\frac{p^{C_2}}{q^d} + e^{-C_3 p}\right] \iint Q\left(\frac{1}{p}, x, k, d\mathbf{y}, d\mathbf{l}\right) = C_1\left[\frac{p^{C_2}}{q^d} + e^{-C_3 p}\right]$$

and the formula (2.63) for $j + 1$ follows. Here $Q(t, x, k, \cdot, \cdot)$ is the transition of probability corresponding to the path measure $\mathfrak{Q}_{x,k}$. $\square$

An estimate of $R_{x,k}^{(\delta)}[B(\delta)]$

We start with a simple observation concerning the Hölder regularity of the K component of any path $\pi \in B(\delta)$. Let us denote $\rho := 2M_\delta^{-1} N^{-1/2}$ and

$$D := \left[\pi \in \mathcal{C}(T,\delta) : |K(t) - K(s)| \geq \rho \text{ for some } k \text{ s.t. } t_k^{(p)} \leq T \text{ and } t_{k-1}^{(p)} \leq s \leq t_k^{(p)} \leq t \leq t_{k+1}^{(p)}\right],$$

where M_δ has been defined in (2.10) and N in (2.27). Suppose that $\pi \in B(\delta)$, then we can find $t \in [t_k^{(p)}, t_{k+1}^{(p)}]$, $s \in [t_{k-1}^{(p)}, t_k^{(p)}]$ for which $\hat{K}(t) \cdot \hat{K}(s) \leq 1 - 1/N$. This, however, implies that

$$|K(t) - K(s)|^2 \geq \frac{1}{M_\delta^2} |\hat{K}(t) - \hat{K}(s)|^2 \geq \frac{2}{M_\delta^2 N},$$

thus $\pi \in D$. Hence the desired estimate of $R_{x,k}^{(\delta)}[B(\delta)]$ follows from the following lemma.

Lemma 2.3.8 *Under the assumptions of Theorem* 2.3.6 *there exist* $C, \gamma > 0$ *such that*

$$R_{x,k}^{(\delta)}[D] \leq CT\delta^\gamma, \quad \forall\, \delta \in (0,1],\ T \geq 1,\ (x,k) \in \mathcal{A}(M). \tag{2.69}$$

Proof. We define the following events:

$$F_1 := \left[|K(t) - K(s)| \geq \rho \quad \text{for some } s,t \in [0,T],\ 0 < t - s < \frac{2}{p},\ t \leq \tau_\delta\right],$$

$$F_2 := \left[|K(t) - K(s)| \geq \rho \quad \text{for some } s,t \in [0,T],\ 0 < t - s < \frac{2}{p},\ s \geq \tau_\delta\right],$$

$$F_3 := \left[|K(\tau_\delta) - K(s)| \geq \frac{\rho}{2} \quad \text{for some } s \in [0,T],\ 0 < \tau_\delta - s < \frac{2}{p},\ \tau_\delta \leq T\right],$$

$$F_4 := \left[|K(\tau_\delta) - K(t)| \geq \frac{\rho}{2} \quad \text{for some } t \in [0,T],\ 0 < t - \tau_\delta < \frac{2}{p}\right].$$

Observe that $D \subset \bigcup_{i=1}^{4} F_i$. Note that F_1, F_3 are $\mathcal{M}^{\tau_\delta}$–measurable, hence

$$R^{(\delta)}_{x,k}[F_i] = \tilde{Q}^{(\delta)}_{x,k}[F_i], \quad i = 1,3. \tag{2.70}$$

On the other hand for $i = 2,4$ we have

$$R^{(\delta)}_{x,k}[F_i] = \int \mathfrak{Q}_{X(\tau_\delta(\pi)),K(\tau_\delta(\pi))}[F_{i,\pi}]\tilde{Q}^{(\delta)}_{x,k}(d\pi),$$

where for a given $\pi \in \mathcal{C}$

$$F_{2,\pi} := \left[|K(t) - K(s)| \geq \rho \quad \text{for some } s,t \in [0,(T-\tau_\delta(\pi)) \wedge 0],\ 0 < t-s < \frac{2}{p}\right],$$

$$F_{4,\pi} := \left[|K(0) - K(t)| \geq \frac{\rho}{2} \quad \text{for some } t \in [0,(T-\tau_\delta(\pi)) \wedge 0],\ 0 < t < \frac{2}{p}\right].$$

Since all F_i, $i = 1,3$ and $F_{i,\pi}$, $i = 2,4$, $\pi \in \mathcal{C}$ are contained in the event

$$F := \left[|K(t) - K(s)| \geq \frac{\rho}{2} \quad \text{for some } s,t \in [0,T],\ 0 < t-s < \frac{2}{p}\right],$$

(2.69) would follow if we show that there exist $C > 0$ and $\gamma > 0$ for which

$$\tilde{Q}^{(\delta)}_{x,k}[F] \leq CT\delta^\gamma \text{ for all } (x,\mathbf{k}) \in \mathcal{A}(M) \tag{2.71}$$

and

$$\mathfrak{Q}_{x,k}[F] \leq CT\delta^\gamma \text{ for all } (x,\mathbf{k}) \in \mathcal{A}(M_\delta). \tag{2.72}$$

The estimate (2.72) follows from elementary properties of diffusions, see e.g. (2.46) p. 47 of [57]. We carry on with the proof of (2.71). The argument is analogous to the proof of Theorem 1.4.6 of [58]. Let L be a multiple of p such that $L := [\delta^{-\gamma_0'}]$, where $\gamma_0' \in (1/2,1)$ is to be specified even further later on. Let also $s^{(L)}_k := k/L$, $k = 0,1,\ldots$. We now define the stopping times $\tau_k(\pi)$ that determine the times at which the K component of the path π performs k–th oscillation of size $\rho/8$. Let $\tau_0(\pi) := 0$ and for any $k \geq 0$

$$\tau_{k+1}(\pi) := \inf\left[s^{(L)}_k \geq \tau_k(\pi) : |K(s^{(L)}_k) - K(\tau_k(\pi))| \geq \frac{\rho}{8}\right],$$

with the convention that $\tau_{n+1} = +\infty$ when $\tau_n = +\infty$, or when the respective event is impossible. Let $N_\# := \min[n : \tau_{n+1} > T]$ and $\delta^* := \min[\tau_n - \tau_{n-1} : n = 1,\ldots,N_\#]$. Then, for a sufficiently small δ_0 and $\delta \in (0,\delta_0)$ we have $F \subset [\delta^* \leq 1/p]$ so we only need to estimate $\tilde{Q}^{(\delta)}_{x,k}$ probability of the latter event.

Let $f : \mathbb{R}^d \to [0,1]$ be a function of $C^\infty_0(\mathbb{R}^d)$ class such that $f(\mathbf{0}) \equiv 1$, when $|k| \leq \rho/16$ and $f(k) \equiv 0$, when $|k| \geq \rho/8$. Let also $f_{\mathbf{l}}(\cdot) := f(\cdot - \mathbf{l})$ for any $\mathbf{l} \in \mathbb{R}^d$. Note that according to Lemma 2.3.5 we can choose constants A_ρ, $C > 0$, where C is independent of ρ, in such a way that $A_\rho < CT^2\rho^{-3}$ and the random sequence

$$S^{\mathbf{l}}_N := \tilde{E}^{(\delta)}_{x,k}\left[f_{\mathbf{l}}\left(K\left(\frac{N+1}{L}\right)\right) \,\middle|\, \mathcal{M}^{N/L}\right] + A_\rho\frac{N}{L}, \quad N \geq 0 \tag{2.73}$$

is a $\tilde{Q}^{(\delta)}_{x,k}$–submartingale with respect to the filtration $\left(\mathcal{M}^{N/L}\right)_{N\geq 0}$ for all $\mathbf{l}$ with the norm $|\mathbf{l}| \in ((3M_\delta)^{-1}, 3M_\delta)$ provided that δ is sufficiently small. We can decompose

$$\tilde{Q}^{(\delta)}_{x,k}\left[\delta^* \leq \frac{2}{p}\right] \leq \tilde{Q}^{(\delta)}_{x,k}\left[\delta^* \leq \frac{2}{p},\ N_\# \leq [\delta^{-\alpha}]\right] + \tilde{Q}^{(\delta)}_{x,k}\left[\delta^* \leq \frac{2}{p},\ N_\# > [\delta^{-\alpha}]\right] \tag{2.74}$$

$$\leq \sum_{i=1}^{[\delta^{-\alpha}]} \tilde{Q}^{(\delta)}_{x,k}\left[\tau_i - \tau_{i-1} \leq \frac{2}{p}\right] + \tilde{Q}^{(\delta)}_{x,k}[N_\# > [\delta^{-\alpha}]],$$

where $\alpha > 0$ is to be determined later. We will show that

$$\tilde{Q}_{x,k}^{(\delta)}[N_{\#} > [\delta^{-\alpha}]] \leq Ce^T \left(1 - \frac{\delta^{1/2(\epsilon_1+\epsilon_2)}}{2}\right)^{\delta^{-\alpha}} \tag{2.75}$$

and

$$\tilde{Q}_{x,k}^{(\delta)}\left[\tau_{n+1} - \tau_n \leq [L\delta_2^{\epsilon}]/L \,\middle|\, \mathcal{M}^{\tau_n}\right] \leq C\delta^{\gamma}T^2, \tag{2.76}$$

for $0 < \gamma < \min[\epsilon_2 - 3\epsilon_1/2, \gamma_0' - (1+\epsilon_1)/2]$. From (2.73), (2.74) (2.75) and (2.76) we further conclude that

$$\tilde{Q}_{x,k}^{(\delta)}\left[\delta^* \leq \frac{1}{p}\right] \leq CT^2\delta^{\gamma-\alpha} + Ce^T\left(1 - \frac{\delta^{1/2(\epsilon_1+\epsilon_2)}}{2}\right)^{\delta^{-\alpha}} \tag{2.77}$$

for some $C > 0$, independent of $\delta \in (0,1]$ and $T \geq T_0$, provided that we choose $\alpha \in (1/2(\epsilon_1 + \epsilon_2), \gamma)$. This is possible if $\min[\epsilon_2 - 3\epsilon_1/2, \gamma_0' - (1+\epsilon_1)/2] > (\epsilon_1+\epsilon_2)/2$, which is true if we assume $\epsilon_2 > 10\epsilon_1 > 0$ and $1 > \gamma_0' > (1+\epsilon_2)/2 + \epsilon_1$. Now, by the argument made after (2.64) we can always replace the first term on the right side of (2.77) by $CT\delta^{\gamma_1}$. We can also assume that the second term on the right hand side of (2.77) is less than or equal to $CT\delta^{\gamma_1}$. This can be seen as follows. Let $\beta := \alpha - 1/2(\epsilon_1+\epsilon_2)$. The term in question is bounded by $C\exp\left\{T - C_1\delta^{-\beta}\right\}$ with $C_1 := \inf_{\rho\in(0,1]}\rho^{-1}\log(1-\rho/2)^{-1}$. For $\delta^{-\beta} \geq 2T/C_1$ we get that $\exp\left\{T - C_1\delta^{-\beta}\right\}$ is less than or equal to $\exp\left\{-C_1\delta^{-\beta}/2\right\}$, while for $\delta^{-\beta} < 2T/C_1$ the left side of (2.77) is obviously less than $2T\delta^{\beta}/C_1$. In both cases we can find a bound as claimed. This proves (2.71) and hence the proof of Lemma 2.3.8 will be complete if we prove (2.75) and (2.76).

To this end, let $\tilde{Q}_{x,k,\pi}^{(\delta)}$, $\pi \in \mathcal{C}$ denote the family of the regular conditional probability distributions that corresponds to $\tilde{Q}_{x,k}^{(\delta)}[\,\cdot \mid \mathcal{M}^{\tau_n}]$. Then, there exists a $\mathcal{M}^{\tau_n}$ measurable, null $\tilde{Q}_{x,k}^{(\delta)}$ probability event Z such that for each $\pi \notin Z$ and each $\mathbf{l} \in \mathbb{R}_*^d$ the random sequence

$$S_{N,\pi}^{\mathbf{l}} := S_N^{\mathbf{l}}\mathbf{1}_{[0,N/L]}(\tau_n(\pi)), \quad N \geq 0$$

is an $\left(\mathcal{M}^{N/L}\right)_{N\geq 0}$ submartingale under $\tilde{Q}_{x,k,\pi}^{(\delta)}$. Let $T_{n,\pi} := \tau_{n+1} \wedge (\tau_n(\pi) + 2[L\delta^{\epsilon}]/L)$, where $\epsilon \in (0,1)$ is a constant to be chosen later on. We can choose the event Z in such a way that

$$\tilde{Q}_{x,k,\pi}^{(\delta)}[T_{n,\pi} \geq \tau_n(\pi)] = 1, \quad \forall\, \pi \notin Z. \tag{2.78}$$

Let $\tilde{S}_{N,\pi} := S_{N,\pi}^{K(\tau_n(\pi))}$, then the submartingale property of $\left(\tilde{S}_{N,\pi}\right)_{N\geq 0}$ and (2.78) imply that

$$\tilde{E}_{x,k,\pi}^{(\delta)}\tilde{S}_{LT_{n,\pi},\pi} \geq \tilde{E}_{x,k,\pi}^{(\delta)}\tilde{S}_{L\tau_n(\pi),\pi} = 1 + A_\rho\tau_n(\pi), \tag{2.79}$$

provided that $\gamma_0 \geq (1+\epsilon_1)/2$. The latter condition assures that $\rho \geq C/(L\sqrt{\delta})$ so that K does not change by more than ρ during the time $1/L$. In consequence of (2.79) we have

$$\tilde{E}_{x,k,\pi}^{(\delta)}\left[f_{K(\tau_n(\pi))}\left(K\left(T_{n,\pi} + \frac{1}{L}\right)\right)\right] + 2A_\rho\delta^{\epsilon} \geq 1, \tag{2.80}$$

as $T_{n,\pi} - \tau_n(\pi) \leq 2[L\delta^{\varepsilon}]/L$. Since

$$\left|f_{K(\tau_n(\pi))}\left(K\left(T_{n,\pi} + \frac{1}{L}\right)\right) - f_{K(\tau_n(\pi))}\left(K\left(T_{n,\pi}\right)\right)\right| \leq \frac{C}{L\rho\delta^{1/2}}$$

we obtain from (2.80)

$$2A_\rho \delta^\epsilon \geq \tilde{E}^{(\delta)}_{x,k,\pi}\left[1 - f_{K(\tau_n(\pi))}\left(K\left(T_{n,\pi}\right)\right)\right] - \frac{C}{L\rho\delta^{1/2}}$$

so in particular

$$2A_\rho \delta^\epsilon + \frac{C}{L\rho\delta^{1/2}} \geq \tilde{E}^{(\delta)}_{x,k,\pi}\left[1 - f_{K(\tau_n(\pi))}\left(K\left(\tau_{n+1}\right)\right), \tau_{n+1} \leq \tau_n(\pi) + \frac{[L\delta^\epsilon]}{L}\right] \quad (2.81)$$
$$= \tilde{Q}^{(\delta)}_{x,k,\pi}\left[\tau_{n+1} \leq \tau_n(\pi) + \frac{[L\delta^\epsilon]}{L}\right].$$

We have shown, therefore, that

$$\tilde{Q}^{(\delta)}_{x,k}\left[\tau_{n+1} - \tau_n \leq \frac{[L\delta^\epsilon]}{L} \,\middle|\, \mathcal{M}^{\tau_n}\right] \leq \frac{CT^2\delta^\epsilon}{\rho^3} + \frac{C}{L\rho\delta^{1/2}} \leq C(\delta^{\epsilon - 3\epsilon_1/2}T^2 + \delta^{\gamma_0' - (1+\epsilon_1)/2}) \leq C\delta^{\gamma_1}T^2 \quad (2.82)$$

for $\gamma_1 < \min[\epsilon - 3\epsilon_1/2, \gamma_0' - (1+\epsilon_1)/2]$ and some constant $C > 0$. We can always assume that $T^2\delta^{\gamma_1/2} \leq 1$. If otherwise, we can always write $\tilde{Q}^{(\delta)}_{x,k}[F] \leq T\delta^{\gamma/4}$ and (2.71) follows. In particular, selecting $\epsilon := (\epsilon_1 + \epsilon_2)/2$, one concludes from (2.82) that

$$\tilde{E}^{(\delta)}_{x,k}[\exp\{-(\tau_{n+1} - \tau_n)\}|\mathcal{M}^{\tau_n}] \leq e^{-\delta^{(\epsilon_1+\epsilon_2)/2}}\tilde{Q}^{(\delta)}_{x,k}\left[\tau_{n+1} - \tau_n \geq \frac{[L\delta^{(\epsilon_1+\epsilon_2)/2}]}{L} \,\middle|\, \mathcal{M}^{\tau_n}\right]$$
$$+\tilde{Q}^{(\delta)}_{x,k}\left[\tau_{n+1} - \tau_n \leq \frac{[L\delta^{(\epsilon_1+\epsilon_2)/2}]}{L} \,\middle|\, \mathcal{M}^{\tau_n}\right] \overset{(2.82)}{\leq} e^{-\delta^{(\epsilon_1+\epsilon_2)/2}} + C\left(1 - e^{-\delta^{(\epsilon_1+\epsilon_2)/2}}\right)\delta^{\gamma/2}$$
$$< 1 - \frac{\delta^{(\epsilon_1+\epsilon_2)/2}}{2} \quad (2.83)$$

provided that δ is sufficiently small. From (2.83) one concludes easily, see e.g. Lemma 1.4.5 p. 38 of [58], that (2.75) holds.

On the other hand, taking $\epsilon = \epsilon_2$ in (2.82) we obtain (2.76) with $0 < \gamma < \min[\epsilon_2 - 3\epsilon_1/2, \gamma_0' - (1+\epsilon_1)/2]$. Hence the proof of Lemma 2.3.8 is now complete. □

2.3.6 The estimation of the convergence rate. The proof of Theorem 2.2.1.

Recall that ϕ_δ, $\bar{\phi}$ satisfy (2.11), (2.13), respectively, with the initial condition ϕ_0. We start with the following lemma.

Lemma 2.3.9 *Assume that ϕ_0 satisfies the hypotheses formulated in Section 2.2.5. Then,*

$$\|\bar{\phi}\|^{[0,T]}_{0,0,0} \leq \|\phi_0\|_{0,0}, \quad \sum_{i=1}^d \|\partial_{x_i}\bar{\phi}\|^{[0,T]}_{0,0,0} \leq \|\phi_0\|_{1,0}. \quad (2.84)$$

Furthermore, there exists a constant $C > 0$ such that for all $T \geq 1$

$$\|\partial_t\bar{\phi}\|^{[0,T]}_{0,0,0} \leq C\|\phi_0\|_{1,2}. \quad (2.85)$$

In addition, for any nonnegative integer valued multi-index $\gamma = (\alpha_1, \alpha_2, \alpha_3)$ satisfying $|\gamma| \leq 3$ we have

$$\sum_{i_1,i_2,i_3=1}^d \|\partial^\gamma_{k_{i_1},k_{i_2},k_{i_3}}\bar{\phi}\|^{[0,T]}_{0,0,0} \leq CT^{|\gamma|}\|\phi_0\|_{1,4}, \quad (2.86)$$

Proof. The estimates (2.84) follow directly from differentiating (2.13) with respect to x. To obtain the estimates (2.85) and (2.86) we note first that the application of the operator $\tilde{\mathcal{L}}$ to both sides of (2.13) and the maximum principle leads to the estimate $\|\tilde{\mathcal{L}}\bar{\phi}(t,x,\cdot)\|_{L^\infty(A(M))} \le \|\tilde{\mathcal{L}}\phi_0\|_{L^\infty(\mathcal{A}(M))}$ for all $t \ge 0$, hence we conclude bound (2.85).

In fact, thanks to already proven estimate (2.84) we conclude that $\|\mathcal{L}\bar{\phi}(t,x,\cdot)\|_{L^\infty(A(M))} \le C\|\phi_0\|_{1,2}$ for some constant $C > 0$ and all $(t,x) \in [0,+\infty) \times \mathbb{R}^d$. Let Z be as in the proof of Lemma 2.3.7. Define $S : Z \times [M^{-1}, M] \to A(M)$ as $S(\mathbf{l},k) := (\mathbf{l}, \sqrt{k^2 - l^2})$, where $l = |\mathbf{l}|$. Let also $\psi(\mathbf{l},k) = \bar{\phi} \circ S(\mathbf{l},k)$. We have $(\mathcal{L}_k\bar{\phi}) \circ S(\mathbf{l},k) = \mathcal{N}\psi(\mathbf{l},k)$, see (2.66). The L^p estimates for elliptic partial differential equations, see e.g. Theorem 9.13 p. 239 of [33] allow us to estimate

$$\|\psi\|_{W^{2,p}(Z)} \le C(\|\psi\|_{L^p(Z)} + \|\mathcal{N}\psi\|_{L^p(Z)}) \le C\|\phi_0\|_{1,2}.$$

Choosing p sufficiently large we obtain that $\sum_i \|\partial_{l_i}\psi\|_{L^\infty(Z)} \le C\|\phi_0\|_{1,2}$, which in fact implies that $\|\mathbf{D}(\cdot)\nabla_k\bar{\phi}(t,\cdot)\|_{L^\infty(S(Z))} \le C\|\phi_0\|_{1,2}$. Obviously, one can find a covering of $A(M)$ with charts corresponding to different choices of the components of k being projected onto the hyperplane $\mathbb{R}^{d-1}$ and we obtain in that way that $\|\mathbf{D}(\cdot)\nabla_k\bar{\phi}(t,\cdot)\|_{L^\infty(\mathcal{A}(M))} \le C\|\phi_0\|_{1,2}$ for all $t \ge 0$. Since the rank of the matrix $\mathbf{D}(\hat{k},k)$ equals $d-1$, with the kernel spanned by the vector k, we obtain in that way the L^∞ estimates of directional derivatives in any direction perpendicular to k. We still need to obtain the L^∞ bound on the derivative in the direction k, denoted by $\partial_n := k_1\partial_{k_1} + \ldots + k_d\partial_{k_d}$. To that purpose we apply ∂_n to both sides of (2.13) and after a straightforward calculation we get $\partial_t\partial_n\bar{\phi} = \tilde{\mathcal{L}}\partial_n\bar{\phi} - 2\mathcal{L}_k\bar{\phi} + \mathcal{L}_1\bar{\phi} + H_0''(k)\hat{k} \cdot \nabla_x\bar{\phi}$, where

$$\mathcal{L}_1\bar{\phi} := \sum_{m,n=1}^d \frac{\partial}{\partial k_m}\left(\partial_k D_{mn}(\hat{k},k)\frac{\partial\bar{\phi}}{\partial k_n}\right).$$

Note that $\mathbf{D}(\hat{k},k)\hat{k} = \mathbf{0}$ implies that $\partial_k\mathbf{D}(\hat{k},k)\hat{k} = \mathbf{0}$ hence $\|\mathcal{L}_1\bar{\phi}(t,\cdot)\|_{L^\infty(\mathcal{A}(M))} \le C\|\phi_0\|_{1,2}$. We already know that $\mathcal{L}_k\bar{\phi}$ and $\|\nabla_x\bar{\phi}\|_{L^\infty(\mathcal{A}(M))}$ are bounded, hence $\|\partial_n\bar{\phi}(t,\cdot)\|_{L^\infty(\mathcal{A}(M))} \le C\|\phi_0\|_{1,2}T$ for $t \in [0,T]$. We have shown therefore that $\|\bar{\phi}(t,\cdot)\|_{1,1} \le C\|\phi_0\|_{1,2}T$ for $t \in [0,T]$. The above procedure can be iterated in order to obtain the estimates of the suprema of derivatives of the higher order. □

Proof of Theorem 2.2.1. Let $u \in [\delta^{\gamma_0'}, T]$, where we assume that γ_0' (as in the statement of Lemma 2.3.5) belongs to the interval $(1/2, 1)$. Substituting for $G(t,x,k) := \bar{\phi}(u-t,x,k)$, $\zeta \equiv 1$ into (2.46) we obtain (taking $v = u$, $t = \delta^{\gamma_0'}$)

$$\left|\tilde{E}_{x,k}^{(\delta)}\left[\phi_0(X(u),K(u)) - \bar{\phi}(u-\delta^{\gamma_0'}, X(\delta^{\gamma_0'}), K(\delta^{\gamma_0'})) - \int_{\delta^{\gamma_0'}}^{u}(\partial_\varrho + \widehat{\mathcal{L}}_\varrho)G(\varrho, X(\varrho), K(\varrho))\,d\varrho\right]\right|$$
$$\le C\|G\|_{1,1,3}^{[0,T]}\delta^{\gamma_1}T^2, \quad \forall\, \delta \in (0,1]. \tag{2.87}$$

Using the fact that $|X(\delta^{\gamma_0'}) - x| \le C\delta^{\gamma_0'}$, $|K(\delta^{\gamma_0'}) - k| \le C\delta^{\gamma_0'-1/2}$, $\tilde{Q}_{x,k}^{(\delta)}$–a.s. for some deterministic constant $C > 0$, cf. (2.36), and Lemma 2.3.9 we obtain that there exist constants $C, \gamma > 0$ such that

$$\left|\tilde{E}_{x,k}^{(\delta)}\left[\phi_0(X(u),K(u)) - \bar{\phi}(u,x,k) - \int_0^u(\partial_\varrho + \widehat{\mathcal{L}}_\varrho)G(\varrho, X(\varrho), K(\varrho))\,d\varrho\right]\right|$$
$$\le C\|G\|_{1,1,3}^{[0,T]}\delta^{\gamma}T^2, \quad \delta \in (0,1],\ T \ge 1,\ u \in [0,T]. \tag{2.88}$$

We have however

$$\left|E_{x,k}^{(\delta)}\left[\phi_0(X(u),K(u))-\bar\phi(u,x,k),\ \tau_\delta\ge T\right]\right| = \left|\tilde E_{x,k}^{(\delta)}\left[\phi_0(X(u),K(u))-\bar\phi(u,x,k),\ \tau_\delta\ge T\right]\right|$$

$$\stackrel{(2.88)}{\le} C\|G\|_{1,1,3}^{[0,T]}\delta^\gamma T^2 + \left(2\|\phi_0\|_{0,0}+T\|G\|_{1,1,2}^{[0,T]}\right)\tilde Q_{x,k}^{(\delta)}[\tau_\delta<T]. \tag{2.89}$$

Using $\mathcal{M}^{\tau_\delta}$ measurability of the event $[\tau_\delta < T]$ we obtain that $\tilde Q_{x,k}^{(\delta)}[\tau_\delta < T] = R_{x,k}^{(\delta)}[\tau_\delta < T]$ and by virtue of Theorem 2.3.6 we can estimate the right hand side of (2.89) by

$$C\|G\|_{1,1,3}^{[0,T]}\delta^\gamma T^2 + C\delta^\gamma T\left(2\|\phi_0\|_{0,0}+T\|G\|_{1,1,2}^{[0,T]}\right) \stackrel{\text{Lemma 2.3.9}}{\le} C\delta^\gamma T^5.$$

On the other hand, the expression under the absolute value on the utmost left hand side of (2.89) equals

$$E_{x,k}^{(\delta)}\left[\phi_0(X(u),K(u))-\bar\phi(u,x,k)\right] - E_{x,k}^{(\delta)}\left[\phi_0(X(u),K(u))-\bar\phi(u,x,k),\ \tau_\delta < T\right].$$

The second term can be estimated by

$$2\|\phi_0\|_{0,0}R_{x,\mathbf{k}}^{(\delta)}[\tau_\delta<T] \stackrel{(2.55)}{\le} C\delta^\gamma\|\phi_0\|_{0,0}T,$$

by virtue of Theorem 2.3.6. Since

$$\mathbb{E}\phi_\delta\left(\frac{u}{\delta},\frac{x}{\delta},k\right) = \mathbb{E}\phi_0(z^{(\delta)}(u;x,k),\mathbf{m}^{(\delta)}(u;x,k)) = E_{x,k}^{(\delta)}\phi_0(X(u),K(u))$$

we conclude from the above that the left hand side of (2.14) can be estimated by $C\delta^\gamma\|\phi_0\|_{1,4}T^5$ for some constants $C,\gamma>0$ independent of $\delta>0$, $T\ge 1$. The bound appearing on the right hand side of (2.14) can be now concluded by the same argument as the one used after (2.64). □

2.4 Momentum diffusion to spatial diffusion: proof of Theorem 2.2.5

We show in this section that solutions of the momentum diffusion equation (2.13) in the long-time, large space limit converge to the solutions of the spatial diffusion equation (2.15). We first recall the setup of Theorem 2.2.5. Let $\bar\phi_\gamma(t,x,k)=\bar\phi(t/\gamma^2,x/\gamma,k)$, where $\bar\phi$ satisfies (2.13) and let $w(t,x,k)$ be the solution of the spatial diffusion equation (2.15). In order to prove Theorem 2.2.5 we need to show that the re-scaled solution $\phi_\gamma(t,x,k)$ converges as $\gamma\to 0$ in the space $C([0,T];L^\infty(\mathcal{A}(M)))$ to $w(t,x,k)$, so that

$$\|w(t)-\bar\phi_\gamma(t)\|_{L^\infty(\mathcal{A}(M))} \le C\left(\gamma T+\gamma^{1/2}\right)\|\phi_0\|_{2,0},\quad 0\le t\le T. \tag{2.90}$$

Proof of Theorem 2.2.5. The proof is quite standard. We present it for the sake of completeness and convenience to the reader. The function $\bar\phi_\gamma$ is the unique $C_b^{1,1,2}([0,+\infty),\mathbb{R}_*^{2d})$-solution to

$$\gamma^2\frac{\partial\bar\phi_\gamma}{\partial t} = \sum_{m,n=1}^d \frac{\partial}{\partial k_m}\left(D_{mn}(\hat k,k)\frac{\partial\bar\phi_\gamma}{\partial k_n}\right) + \gamma H_0'(k)\hat k\cdot\nabla_x\bar\phi_\gamma. \tag{2.91}$$

$$\bar\phi_\gamma(0,x,k)=\phi_0(x,k),$$

see Remark 2.2.3. We represent $\bar{\phi}_\gamma$ as

$$\bar{\phi}_\gamma = w + \gamma w_1 + \gamma^2 w_2 + R. \tag{2.92}$$

Here w is the solution of the diffusion equation (2.15), the correctors w_1 and w_2 will be constructed explicitly, and the remainder R will be shown to be small. The first corrector w_1 is the unique solution of zero mean over each sphere $\mathbb{S}_k^{d-1}$ of the equation

$$\sum_{m,n=1}^{d} \frac{\partial}{\partial k_m}\left(D_{mn}(\hat{k},k)\frac{\partial w_1}{\partial k_n}\right) = -H_0'(k)\hat{k}\cdot\nabla_x w. \tag{2.93}$$

It has an explicit form

$$w_1(t,x,k) = \sum_{j=1}^{d} \chi_j(k)\frac{\partial w(t,x,k)}{\partial x_j} \tag{2.94}$$

with the functions χ_j defined in (2.17). The second order corrector w_2 is the unique zero mean over each sphere $\mathbb{S}_k^{d-1}$ solution of the equation

$$\sum_{m,n=1}^{d} \frac{\partial}{\partial k_m}\left(D_{mn}(\hat{k},k)\frac{\partial w_2}{\partial k_n}\right) = \frac{\partial w}{\partial t} - H_0'(k)\hat{k}\cdot\nabla_x w_1. \tag{2.95}$$

Note that the expression on the right hand side of (2.95) is of zero mean since thanks to (2.15) and equality (2.16) we have

$$\frac{\partial w}{\partial t} = \frac{1}{\Gamma_{d-1}}\int_{\mathbb{S}^{d-1}} H_0'(k)\hat{k}\cdot\nabla_x w_1 d\Omega(\hat{k}).$$

Equations (2.93) and (2.95) for various values of $k = |k|$ are decoupled. As a consequence of this fact and the regularity properties for solutions of elliptic equations on a sphere we have that w_1, w_2 belong to $C([0,T];L^\infty(\mathcal{A}(M)))$. More explicitly, we may represent the function w_2 as

$$w_2(t,x,k) = \sum_{j,l=1}^{d} \psi_{jl}(k)\frac{\partial^2 w(t,x,k)}{\partial x_j \partial x_l}.$$

The functions $\psi_{jm}(k)$ satisfy

$$\sum_{m,n=1}^{d} \frac{\partial}{\partial k_m}\left(D_{mn}(\hat{k},k)\frac{\partial \psi_{jl}}{\partial k_n}\right) = -H_0'(k)\hat{k}_j\chi_l(k) + a_{jl}(k). \tag{2.96}$$

A unique mean-zero, bounded solution of (2.96) exists according to the Fredholm alternative combined the regularity properties for solutions of (2.96) on each sphere $\mathbb{S}_k^{d-1}$. With the above definitions of w, w_1, w_2, equation (2.91) for $\bar{\phi}_\gamma$ implies that the remainder R in (2.92) satisfies

$$\gamma^2\frac{\partial R}{\partial t} + \gamma^3\frac{\partial w_1}{\partial t} + \gamma^4\frac{\partial w_2}{\partial t} - \gamma H_0'(k)\hat{k}\cdot\nabla_x R - \gamma^3 H_0'(k)\hat{k}\cdot\nabla_x w_2 = \sum_{m,n=1}^{d}\frac{\partial}{\partial k_m}\left(D_{mn}(\hat{k},k)\frac{\partial R}{\partial k_n}\right).$$

We re-write this equation in the form

$$\frac{\partial R}{\partial t} - \frac{1}{\gamma}H_0'(k)\hat{k}\cdot\nabla_x R - \frac{1}{\gamma^2}\sum_{m,n=1}^{d}\frac{\partial}{\partial k_m}\left(D_{mn}(\hat{k},k)\frac{\partial R}{\partial k_n}\right) = f \tag{2.97}$$

$$R(0,x,k) = \phi_0(x,k) - \bar{\phi}_0(x,k) - \gamma w_1(0,x,k) - \gamma^2 w_2(0,x,k),$$

where $f := -\gamma\partial_t w_1 - \gamma^2\partial_t w_2 - \gamma H_0'(k)\hat{k}\cdot\nabla_x w_2$. Here, as before, R is understood as the unique solution to (2.97) that belongs to $C_b^{1,1,2}([0,+\infty),\mathbb{R}_*^{2d})$. We may split $R = R_1 + R_2$ according to the initial data and forcing in the equation: R_1 satisfies

$$\frac{\partial R_1}{\partial t} - \frac{1}{\gamma}H_0'(k)\hat{k}\cdot\nabla_x R_1 - \frac{1}{\gamma^2}\sum_{m,n=1}^{d}\frac{\partial}{\partial k_m}\left(D_{mn}(\hat{k},k)\frac{\partial R_1}{\partial k_n}\right) = f,$$
$$R(0,x,k) = -\gamma w_1(0,x,k) - \gamma^2 w_2(0,x,k) \tag{2.98}$$

and the initial time boundary layer corrector R_2 satisfies

$$\frac{\partial R_2}{\partial t} - \frac{1}{\gamma}H_0'(k)\hat{k}\cdot\nabla_x R_2 - \frac{1}{\gamma^2}\sum_{m,n=1}^{d}\frac{\partial}{\partial k_m}\left(D_{mn}(\hat{k},k)\frac{\partial R_2}{\partial k_n}\right) = 0 \tag{2.99}$$
$$R_2(0,x,k) = \phi_0(x,k) - \bar{\phi}_0(x,k).$$

Using the probabilistic representation for the solution of (2.98) as well as the regularity of w_1 and w_2 we obtain that

$$\|R_1(t)\|_{L^\infty(\mathcal{A}(M))} \le C\gamma T, \quad 0 \le t \le T. \tag{2.100}$$

To obtain the bound for R_2 we consider $R_2^\gamma(t,x,k) := R_2(\gamma^{3/2}t,x,k)$. This function satisfies

$$\frac{\partial R_2^\gamma}{\partial t} - \gamma^{1/2}H_0'(k)\hat{k}\cdot\nabla_x R_2^\gamma - \frac{1}{\gamma^{1/2}}\sum_{m,n=1}^{d}\frac{\partial}{\partial k_m}\left(D_{mn}(\hat{k},k)\frac{\partial R_2^\gamma}{\partial k_n}\right) = 0$$
$$R_2^\gamma(0,x,k) = \phi_0(x,k) - \bar{\phi}_0(x,k).$$

We also define $\tilde{R}_2^\gamma$, the solution of

$$\frac{\partial \tilde{R}_2^\gamma}{\partial t} - \frac{1}{\gamma^{1/2}}\sum_{m,n=1}^{d}\frac{\partial}{\partial k_m}\left(D_{mn}(\hat{k},k)\frac{\partial \tilde{R}_2^\gamma}{\partial k_n}\right) = 0 \tag{2.101}$$
$$\tilde{R}_2^\gamma(0,x,k) = \phi_0(x,k) - \bar{\phi}_0(x,k).$$

The uniform ellipticity of the right hand side of (2.101) on each sphere $\mathbb{S}_k^{d-1}$ implies, see e.g. Proposition 13.3, p. 55 of [60] that the function $\tilde{R}_2^\gamma$ satisfies the decay estimate on each sphere

$$\|\tilde{R}_2^\gamma(t)\|_{L^\infty(\mathbb{S}^{d-1})} \le \frac{C\gamma^{(d-1)/4}}{t^{(d-1)/2}}\|\phi_0\|_{L^1(\mathbb{S}_k^{d-1})} \le \frac{C\gamma^{(d-1)/4}}{t^{(d-1)/2}}\|\phi_0\|_{L^\infty(\mathbb{S}_k^{d-1})} \tag{2.102}$$

for $t \in [0,T]$ and, similarly,

$$\|\nabla_x \tilde{R}_2^\gamma(t)\|_{L^\infty(\mathbb{S}_k^{d-1})} \le \frac{C\gamma^{(d-1)/4}}{t^{(d-1)/2}}\|\phi_0\|_{1,0}.$$

Furthermore, the difference $q^\gamma = R_2^\gamma - \tilde{R}_2^\gamma$ satisfies

$$\frac{\partial q^\gamma}{\partial t} - \gamma^{1/2}H_0'(k)\hat{k}\cdot\nabla_x q^\gamma - \frac{1}{\gamma^{1/2}}\sum_{m,n=1}^{d}\frac{\partial}{\partial k_m}\left(D_{mn}(\hat{k},k)\frac{\partial q^\gamma}{\partial k_n}\right) = \gamma^{1/2}H_0'(k)\hat{k}\cdot\nabla_x \tilde{R}_2^\gamma \tag{2.103}$$
$$q^\gamma(0,x,k) = 0.$$

We conclude, using the probabilistic representation of the solution of (2.103), that

$$\|q^\gamma(t)\|_{L^\infty(\mathcal{A}(M))} \le C\gamma^{1/2}t\|\phi_0\|_{1,0}$$

and thus

$$\|R_2(\gamma^{3/2})\|_{L^\infty(\mathcal{A}(M))} \le \|R_2^\gamma(1)\|_{L^\infty(\mathcal{A}(M))} + \|q^\gamma(1)\|_{L^\infty(\mathcal{A}(M))}$$
$$\le C\left(\gamma^{(d-1)/4}\|\phi_0\|_{0,0} + \gamma^2\|\phi_0\|_{1,0}\right).$$

The maximum principle for (2.99) implies that we have the above estimate for all $t \ge \gamma^{3/2}$:

$$\|R_2(t)\|_{L^\infty(\mathcal{A}(M))} \le C\left(\gamma^{(d-1)/4}\|\phi_0\|_{0,0} + \gamma^{1/2}\|\nabla_x\phi_0\|_{1,0}\right), \quad t \ge \gamma^{3/2}. \tag{2.104}$$

Combining (2.92), (2.100) and (2.104) we conclude that

$$\|w(t) - \bar{\phi}_\gamma(t)\|_{L^\infty(\mathcal{A}(M))} \le C\left(\gamma T + \gamma^{(d-1)/4} + \gamma^{1/2}\right)\|\phi_0\|_{1,0}, \quad \gamma^{3/2} \le t \le T, \tag{2.105}$$

and thus (2.90) follows, as $d \ge 3$. This finishes the proof of Theorem 2.2.5. □

2.5 The spatial diffusion of wave energy

In this section we consider an application of the previous results to the random geometrical optics regime of propagation of acoustic waves. We show that when the wave length is much shorter than the correlation length of the random medium, there exist temporal and spatial scales where the energy density of the wave undergoes the spatial diffusion. We start with the wave equation in dimension $d \ge 3$

$$\frac{1}{c^2(x)}\frac{\partial^2\phi}{\partial t^2} - \Delta\phi = 0 \tag{2.106}$$

and assume that the wave speed has the form $c(x) = c_0 + \sqrt{\delta}c_1(x)$. Here $c_0 > 0$ is the constant sound speed of the uniform background medium, while the small parameter $\delta \ll 1$ measures the strength of the mean zero random perturbation c_1. Rescaling the spatial and temporal variables $x = x'/\delta$ and $t = t'/\delta$ we obtain (after dropping the primes) equation (2.106) with a rapidly fluctuating wave speed

$$c_\delta(x) = c_0 + \sqrt{\delta}c_1\left(\frac{x}{\delta}\right). \tag{2.107}$$

It is convenient to rewrite (2.106) as the system of acoustic equations for the "pressure" $p = \phi_t/c$ and "acoustic velocity" $\mathbf{u} = -\nabla\phi$:

$$\begin{aligned}&\frac{\partial\mathbf{u}}{\partial t} + \nabla\left(c_\delta(x)p\right) = 0\\&\frac{\partial p}{\partial t} + c_\delta(x)\nabla\cdot\mathbf{u} = 0.\end{aligned} \tag{2.108}$$

We will denote for brevity $\mathbf{v} = (\mathbf{u}, p) \in \mathbb{R}^{d+1}$ and write (2.108) in the more general form of a first order linear symmetric hyperbolic system. To do so we introduce symmetric matrices A_δ and D^j defined by

$$A_\delta(x) = \mathrm{diag}(1,1,1,c_\delta(x)), \qquad \text{and} \qquad D^j = \mathbf{e}_j \otimes \mathbf{e}_{d+1} + \mathbf{e}_{d+1}\otimes\mathbf{e}_j,\ j = 1,\dots,d. \tag{2.109}$$

Here $\mathbf{e}_m \in \mathbb{R}^{d+1}$ is the standard orthonormal basis: $(\mathbf{e}_m)_k = \delta_{mk}$.

We consider the initial data for (2.108) as a mixture of states. Let S be a measure space equipped with a non-negative finite measure μ. A typical example is that the initial data is random, S is the state space and μ is the corresponding probability measure. We assume that for each parameter $\zeta \in S$ and $\varepsilon, \delta > 0$ the initial data is given by $\mathbf{v}_\varepsilon^\delta(0, x; \zeta) := (-\varepsilon\nabla\phi_0^\varepsilon(x), 1/c_\delta(x)\dot{\phi}_0^\varepsilon(x))$ and $\mathbf{v}_\varepsilon^\delta(t, x; \zeta)$ solves the system of equations

$$\frac{\partial \mathbf{v}_\varepsilon^\delta}{\partial t} + \sum_{j=1}^{d} A_\delta(x) D^j \frac{\partial}{\partial x^j}\left(A_\delta(x)\mathbf{v}_\varepsilon^\delta(x)\right) = 0. \tag{2.110}$$

The set of initial data is assumed to form an ε-oscillatory and compact at infinity family [31] as $\varepsilon \to 0$. By the above we mean that for any continuous, compactly supported function $\varphi : \mathbb{R}^d \to \mathbb{R}$ we have

$$\lim_{R\to+\infty} \limsup_{\varepsilon\to 0^+} \int_{|k|\geq R/\varepsilon} |\widehat{\varphi \mathbf{v}_\varepsilon^\delta}|^2 dk \to 0 \text{ and } \lim_{R\to+\infty} \limsup_{\varepsilon\to 0^+} \int_{|x|\geq R} |\mathbf{v}_\varepsilon^\delta|^2 dx \to 0$$

for a fixed realization $\zeta \in S$ of the initial data and each $\delta > 0$. In the regime of geometric acoustics the scale ε of oscillations of the initial data is much smaller than the correlation length δ of the medium: $\varepsilon \ll \delta \ll 1$.

The dispersion matrix for (2.110) is

$$P_0^\delta(x,k) = i\sum_{j=1}^{d} A_\delta(x) k_j D^j A_\delta(x) = i\sum_{j=1}^{d} c_\delta(x) k_j D^j = ic_\delta(x)\left(\tilde{k}\otimes \mathbf{e}_{d+1} + \mathbf{e}_{d+1}\otimes\tilde{k}\right), \tag{2.111}$$

where $\tilde{k} = \sum_{j=1}^{d} k_j \mathbf{e}_j$. The self-adjoint matrix $(-iP_0^\delta)$ has an eigenvalue $H_0 = 0$ of the multiplicity $d-1$, and two simple eigenvalues

$$H_\pm^\delta(x,k) = \pm c_\delta(x)|k|. \tag{2.112}$$

Its eigenvectors are

$$b_m^0 = \left(k_m^\perp, 0\right),\; m = 1, \ldots, d-1;\;\; b_\pm = \frac{1}{\sqrt{2}}\left(\frac{\tilde{k}}{|k|} \pm \mathbf{e}_{d+1}\right), \tag{2.113}$$

where $k_m^\perp \in \mathbb{R}^d$ is the orthonormal basis of vectors orthogonal to k.

The $(d+1)\times(d+1)$ Wigner matrix of a mixture of solutions of (2.110) is defined by

$$W_\varepsilon^\delta(t,x,k) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\int_S e^{ik\cdot y}\mathbf{v}_\varepsilon^\delta(t, x - \frac{\varepsilon y}{2};\zeta)\mathbf{v}_\varepsilon^{\delta *}(t, x + \frac{\varepsilon y}{2};\zeta) dy\mu(d\zeta). \tag{2.114}$$

It is well-known, see [31, 49, 53], that for each fixed $\delta > 0$ (and even without introduction of a mixture of states) when $W_\varepsilon^\delta(t=0)$ converges weakly in $\mathcal{S}'(\mathbb{R}^d\times\mathbb{R}^d)$, as $\varepsilon \to 0$, to

$$W_0(x,k) = u_+^0(x,k) b_+(k)\otimes b_+(k) + u_-^0(x,k) b_-(k)\otimes b_-(k). \tag{2.115}$$

then $W_\varepsilon^\delta(t)$ converges weakly in $\mathcal{S}'(\mathbb{R}^d\times\mathbb{R}^d)$ to

$$U^\delta(t,x,k) = u_+^\delta(t,x,k) b_+(k)\otimes b_+(k) + u_-^\delta(t,x,k) b_-(k)\otimes b_-(k).$$

The scalar amplitudes $u_\pm^{(\delta)}$ satisfy the Liouville equations:

$$\left\{\begin{array}{l} \partial_t u_\pm^\delta + \nabla_k H_\pm^\delta \cdot \nabla_x u_\pm^\delta - \nabla_x H_\pm^\delta \cdot \nabla_k u_\pm^\delta = 0, \\ \\ u_\pm^\delta(0,x,k) = u_\pm^0(x,k). \end{array}\right. \tag{2.116}$$

These equations are of the form (2.11), written in the macroscopic variables, with the Hamiltonian given by (2.112).

One may obtain an L^2-error estimate for this convergence when a mixture of states is introduced, as in (2.114). In order to make the scale separation $\varepsilon \ll \delta \ll 1$ precise we define the set

$$\mathcal{K}_\mu := \left\{(\varepsilon,\delta): \quad |\ln\varepsilon|^{-2/3+\mu} \le \delta \le 1\right\}.$$

The parameter μ is a fixed number in the interval $(0, 2/3)$. The following proposition has been proved in Theorem 3.2 of [6], using straightforward if tedious asymptotic expansions.

Proposition 2.5.1 *Let the acoustic speed $c_\delta(x)$ be of the form (2.107) and such that the Hamiltonian $H_\delta(x)$ given by (2.112) satisfies assumptions (2.6). We assume that the Wigner transform W_ε^δ satisfies*

$$W_\varepsilon^\delta(0,x,k) \to W_0(x,k) \text{ strongly in } L^2(\mathbb{R}^d \times \mathbb{R}^d) \text{ as } \mathcal{K}_\mu \ni (\varepsilon,\delta) \to 0. \tag{2.117}$$

We also assume that the limit $W_0 \in C_c^2(\mathbb{R}^d \times \mathbb{R}^d)$ with a support that satisfies

$$\text{supp } W_0(x,k) \subseteq \mathcal{A}(M) \tag{2.118}$$

for some $M > 0$. Moreover, we assume that the initial limit Wigner transform W_0 is of the form

$$W_0(x,k) = \sum_{q=\pm} u_q^0(x,k)\Pi_q(k), \quad \Pi_q(k) = b_q(k) \otimes b_q(k). \tag{2.119}$$

Let $U^\delta(t,x,k) = \sum_{p=\pm} u_p^\delta(t,x,k)\Pi_p(k)$, where the functions u_p^δ satisfy the Liouville equations (2.116). Then there exists a constant $C_1 > 0$ that is independent of δ so that

$$\|W_\varepsilon^\delta(t,x,k) - U^\delta(t,x,k)\|_2 \le C(\delta)\left(\varepsilon\|W_0\|_{H^2}e^{C_1 t/\delta^{3/2}} + \varepsilon^2\|W_0\|_{H^3}e^{C_1 t/\delta^{3/2}}\right) + \|W_\varepsilon^\delta(0) - W_0\|_2, \tag{2.120}$$

where $C(\delta)$ is a rational function of δ with deterministic coefficients that may depend on the constant $M > 0$ in the bound (2.118) on the support of W_0.

The Liouville equations (2.116) are of the form (2.11). Therefore, one may pass to the limit $\delta \to 0$ in (2.116) using Theorem 2.2.1 and conclude that $\mathbb{E}u_\pm^\delta$ converge to the respective solutions of

$$\frac{\partial \bar{u}_\pm}{\partial t} = \sum_{m,n=1}^d \frac{\partial}{\partial k_m}\left(|k|^2 D_{mn}(\hat{k})\frac{\partial \bar{u}_\pm}{\partial k_n}\right) \pm c_0\hat{k}\cdot\nabla_x\bar{u}_\pm \tag{2.121}$$

with the initial conditions as in (2.116). Here the diffusion matrix $D(\hat{k}) = [D_{mn}(\hat{k})]$ is given by

$$D_{mn}(\hat{k}) = -\frac{1}{2}\int_{-\infty}^{\infty} \frac{\partial^2 R(c_0 s\hat{k})}{\partial x_n \partial x_m} ds, \tag{2.122}$$

where $R(x)$ is the correlation function of the random field $c_1(x)$: $\mathbb{E}\left[c_1(z)c_1(x+z)\right] = R(x)$. Furthermore, it follows from Theorem 2.2.7 that there exists $\alpha_0 > 0$ so that solutions of (2.116) with the initial data of the form $u_\pm^\delta(0,x,k) = u_\pm^0(\delta^\alpha x, k)$ with $0 < \alpha < \alpha_0$, converge in the long time limit to the solutions of the spatial diffusion equation. More precisely, in that case the function $\bar{u}^\delta(t,x,k) = u_+^\delta(t/\delta^{2\alpha}, x/\delta^\alpha, k)$ (and similarly for u_-^δ) converges as $\delta \to 0$ to $w(t,x,k)$ – the solution of the spatial diffusion equation

$$\frac{\partial w}{\partial t} = \sum_{m,n=1}^{d} a_{mn}(k) \frac{\partial^2 w}{\partial x_n \partial x_m}, \tag{2.123}$$
$$w(0,x,k) = \bar{u}_+^0(x;k) := \frac{1}{\Gamma_{d-1}} \int_{\mathbb{S}^{d-1}} u_+^0(x,k) d\Omega(\hat{k}).$$

with the diffusion matrix a_{mn} given by:

$$a_{nm}(k) = \frac{c_0}{\Gamma_{d-1}} \int_{\mathbb{S}^{d-1}} \hat{k}_n \chi_m(k) d\Omega(\hat{k}), \tag{2.124}$$

and the functions χ_j above are the mean-zero solutions of

$$\sum_{m,n=1}^{d} \frac{\partial}{\partial k_m} \left(k^2 D_{mn}(\hat{k}) \frac{\partial \chi_j}{\partial k_n} \right) = -c_0 \hat{k}_j. \tag{2.125}$$

Theorems 2.2.1, 2.2.5 and 2.2.7 allow us to make the passage to the limit $\varepsilon, \delta, \gamma \to 0$ rigorous. The assumption that $\varepsilon \ll \delta \ll \gamma$ is formalized as follows. We let

$$\mathcal{K}_{\mu,\rho} := \left\{ (\varepsilon, \delta, \gamma) : \quad \delta \geq |\ln \varepsilon|^{-2/3+\mu} \text{ and } \gamma \geq \delta^\rho \right\},$$

with $0 < \mu < 2/3$, $\rho \in (0,1)$. Suppose also that $u_0^\pm \in C_c^3(\mathbb{R}_*^{2d})$ and supp $u_0^\pm \subseteq \mathcal{A}(M)$. Let

$$W^0(x,k) := u_+^0(x,k) b_+(k) \otimes b_+(k) + u_-^0(x,k) b_-(k) \otimes b_-(k), \tag{2.126}$$

and

$$W(t,x,k) := w_+(t,x;k) b_+(k) \otimes b_+(k) + w_-(t,x;k) b_-(k) \otimes b_-(k). \tag{2.127}$$

Our main result regarding the diffusion of wave energy can be stated as follows.

Theorem 2.5.2 *Assume that the dimension $d \geq 3$ and $M \geq 1$ are fixed. Suppose for some $0 < \mu < 2/3$, $\rho \in (0,1)$ we have, with W^0 as in* (2.126) *and W_ε^δ defined by* (2.114)

$$\int\limits_{\mathbb{R}^{2d}} \left| \mathbb{E} W_\varepsilon^\delta \left(0, \frac{x}{\gamma}, k \right) - W^0(x,k) \right|^2 dx dk \to 0, \text{ as } (\varepsilon, \delta, \gamma) \to 0 \text{ and } (\varepsilon, \delta, \gamma) \in \mathcal{K}_{\mu,\rho}.$$

Then, there exists $\rho_1 \in (0, \rho]$ such that for any $T > T_* > 0$ we have

$$\sup_{t \in [T_*, T]} \int \left| \mathbb{E} W_\varepsilon^\delta \left(\frac{t}{\gamma^2}, \frac{x}{\gamma}, k \right) - W(t,x,k) \right|^2 dx dk \to 0, \text{ as } (\varepsilon, \delta, \gamma) \to 0 \text{ and } (\varepsilon, \delta, \gamma) \in \mathcal{K}_{\mu,\rho_1}.$$

Here $W(t,x,k)$ is of the form (2.127) *with the functions $w_\pm$ that satisfy* (2.123) *with the initial data $w_\pm(0,x,k) = u_\pm^0(x,k)$.*

The proof follows immediately from Theorems 2.2.1, 2.2.5 and 2.2.7 as well as Proposition 2.5.1.

2.6 The proof of Lemma 2.3.5.

Given $s \geq \sigma > 0$, $\pi \in \mathcal{C}$ we define the linear approximation of the trajectory

$$\boldsymbol{L}(\sigma, s; \pi) := X(\sigma) + (s-\sigma)H_0'(K(\sigma))\hat{K}(\sigma) \tag{2.128}$$

and for any $v \in [0,1]$ let

$$\boldsymbol{R}(v, \sigma, s; \pi) := (1-v)\boldsymbol{L}(\sigma, s; \pi) + vX(s). \tag{2.129}$$

The following simple fact can be verified by a direct calculation, see Lemma 5.4 of [6].

Proposition 2.6.1 *Suppose that $s \geq \sigma \geq 0$ and $\pi \in \mathcal{C}(\delta)$. Then,*

$$|X(s) - \boldsymbol{L}(\sigma, s; \pi)| \leq \tilde{D}(2M_\delta)\sqrt{\delta}(s-\sigma) + \int_\sigma^s |H_0'(K(\rho))\hat{K}(\rho) - H_0'(K(\sigma))\hat{K}(\sigma)| d\rho.$$

We obtain from Proposition 2.6.1 for each $s \geq \sigma$ an error for the first-order approximation of the trajectory

$$|z^{(\delta)}(s) - \boldsymbol{l}^{(\delta)}(\sigma, s)| \leq \tilde{D}(2M_\delta)\sqrt{\delta}(s-\sigma) + \frac{C(s-\sigma)^2}{2\sqrt{\delta}}, \quad \delta \in (0, \delta_*(M)].$$

Here $\boldsymbol{l}^{(\delta)}(\sigma, s) := z^{(\delta)}(\sigma) + (s-\sigma)\hat{\mathbf{m}}^{(\delta)}(\sigma)$ is the linear approximation between the times σ and s and

$$C := \sup_{\delta \in (0, \delta_*(M)]} (M_\delta h_0^*(M_\delta) + \tilde{h}_0^*(M_\delta))\tilde{D}(2M_\delta).$$

With no loss of generality we may assume that $\mathbf{x} = 0$ and that there exists k such that $t, u \in [t_k^{(p)}, t_{k+1}^{(p)})$. We shall omit the initial condition in the notation of the solution to (2.36). Throughout this argument we use Proposition 2.6.1 with

$$\sigma(s) := s - \delta^{1-\gamma_A} \text{ for some } \gamma_A \in (0, 1/16 \wedge (1-\epsilon_4)), \quad s \in [t, u]. \tag{2.130}$$

The aforementioned proposition proves that for this choice of σ we have

$$|\boldsymbol{L}^{(\delta)}(\sigma, s) - y^{(\delta)}(s)| \leq C_A \delta^{3/2 - 2\gamma_A}, \quad \forall\, \delta \in (0, 1]. \tag{2.131}$$

Throughout this section we denote $\tilde{\zeta} = \zeta(y^{(\delta)}(t_1), \boldsymbol{l}^{(\delta)}(t_1), \ldots, y^{(\delta)}(t_n), \boldsymbol{l}^{(\delta)}(t_n))$. We assume first that $G \in C^2(\mathbb{R}_*^d)$ and $\|G\|_2 < +\infty$. Note that

$$G(\boldsymbol{l}^{(\delta)}(u)) - G(\boldsymbol{l}^{(\delta)}(t)) = -\frac{1}{\sqrt{\delta}} \sum_{j=1}^d \int_t^u \partial_j G(\boldsymbol{l}^{(\delta)}(s)) F_{j,\delta}\left(s, \frac{y^{(\delta)}(s)}{\delta}, \boldsymbol{l}^{(\delta)}(s)\right) ds. \tag{2.132}$$

We can rewrite then (2.132) in the form $I^{(1)} + I^{(2)} + I^{(3)}$, where

$$I^{(1)} := -\frac{1}{\sqrt{\delta}} \sum_{j=1}^d \int_t^u \partial_j G(\boldsymbol{l}^{(\delta)}(\sigma)) F_{j,\delta}\left(s, \frac{y^{(\delta)}(s)}{\delta}, \boldsymbol{l}^{(\delta)}(\sigma)\right) ds,$$

$$I^{(2)} := \frac{1}{\delta} \sum_{i,j=1}^d \int_t^u \int_\sigma^s \partial_j G(\boldsymbol{l}^{(\delta)}(\rho)) \partial_{\ell_i} F_{j,\delta}\left(s, \frac{y^{(\delta)}(s)}{\delta}, \boldsymbol{l}^{(\delta)}(\rho)\right) F_{i,\delta}\left(\rho, \frac{y^{(\delta)}(\rho)}{\delta}, \boldsymbol{l}^{(\delta)}(\rho)\right) ds\, d\rho,$$

$$I^{(3)} := \frac{1}{\delta} \sum_{i,j=1}^d \int_t^u \int_\sigma^s \partial_{i,j}^2 G(\boldsymbol{l}^{(\delta)}(\rho)) F_{j,\delta}\left(s, \frac{y^{(\delta)}(s)}{\delta}, \boldsymbol{l}^{(\delta)}(\rho)\right) F_{i,\delta}\left(\rho, \frac{y^{(\delta)}(\rho)}{\delta}, \boldsymbol{l}^{(\delta)}(\rho)\right) ds\, d\rho$$

and σ is given by (2.130). Each of these terms will be estimated separately below.

The term $\mathbb{E}[I^{(1)}\tilde{\zeta}]$

The term $I^{(1)}$ can be rewritten in the form $J^{(1)} + J^{(2)}$, where

$$J^{(1)} := -\frac{1}{\sqrt{\delta}} \sum_{j=1}^{d} \int_t^u \partial_j G(\boldsymbol{l}^{(\delta)}(\sigma)) F_{j,\delta} \left(s, \frac{\boldsymbol{L}^{(\delta)}(\sigma, s)}{\delta}, \boldsymbol{l}^{(\delta)}(\sigma) \right) ds,$$

and

$$J^{(2)} := -\frac{1}{\delta^{3/2}} \sum_{i,j=1}^{d} \int_t^u \int_0^1 \partial_j G(\boldsymbol{l}^{(\delta)}(\sigma)) \partial_{y_i} F_{j,\delta} \left(s, \frac{\boldsymbol{R}^{(\delta)}(v, \sigma, s)}{\delta}, \boldsymbol{l}^{(\delta)}(\sigma) \right) (y_i^{(\delta)}(s) - L_i^{(\delta)}(\sigma, s))\, ds\, dv, \tag{2.133}$$

where (see (2.128) and (2.129) for the definition of $\boldsymbol{L}$ and $\boldsymbol{R}$) we abbreviate $\boldsymbol{L}^{(\delta)}(\sigma, s) = \boldsymbol{L}(\sigma, s; y^{(\delta)}(\cdot), \boldsymbol{l}^{(\delta)}(\cdot))$, $\boldsymbol{R}^{(\delta)}(\sigma, s) = \boldsymbol{R}(\sigma, s; y^{(\delta)}(\cdot), \boldsymbol{l}^{(\delta)}(\cdot))$. We use part (i) of Lemma 2.3.2 to handle the term $\mathbb{E}[J^{(1)}\tilde{\zeta}]$. Let $\tilde{X}_1(\mathbf{x}, k) = -\partial_{x_i} H_1(\mathbf{x}, k)$, $\tilde{X}_2(\mathbf{x}, k) \equiv 1$,

$$Z = \Theta \left(t_k^{(p)}, \frac{\boldsymbol{L}^{(\delta)}(\sigma, s)}{\delta}, \boldsymbol{l}^{(\delta)}(\sigma) \right) \partial_j G(\boldsymbol{l}^{(\delta)}(\sigma)) \tilde{\zeta}$$

and $g_1 = (\boldsymbol{L}^{(\delta)}(\sigma, s)\delta^{-1}, |\boldsymbol{l}^{(\delta)}(\sigma)|)$. Note that g_1 and Z are both $\mathcal{F}_\sigma$ measurable. We need to verify (2.39). Suppose therefore that $Z \neq 0$. For $\rho \in [0, t_{k-1}^{(p)}]$ we have $|\boldsymbol{L}^{(\delta)}(\sigma, s) - y^{(\delta)}(\rho)| \geq (2q)^{-1}$, provided that $C_A \delta^{3/2 - 2\gamma_A} < 1/(2q)$, which holds for sufficiently small δ, since our assumptions on the exponents $\epsilon_2, \epsilon_3, \gamma_A$ (namely that $\epsilon_2, \epsilon_3 < 1/8$, $\gamma_A < 1/8$) guarantee that $\epsilon_2 + \epsilon_3 < 3/4 - \gamma_A/2$. For $\rho \in [t_{k-1}^{(p)}, \sigma]$ we have

$$\begin{aligned}
(\boldsymbol{L}^{(\delta)}(\sigma, s) - y^{(\delta)}(\rho)) \cdot \hat{\boldsymbol{l}}^{(\delta)} \left(t_{k-1}^{(p)} \right) &\geq (s - \sigma) H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)\, \hat{\boldsymbol{l}}^{(\delta)}(\sigma) \cdot \hat{\boldsymbol{l}}^{(\delta)} \left(t_{k-1}^{(p)} \right) \\
&+ \int_\rho^\sigma \left[H_0'(|\boldsymbol{l}^{(\delta)}(\rho_1)|) + \sqrt{\delta}\, \partial_l H_1 \left(\frac{y^{(\delta)}(\rho_1)}{\delta}, |\boldsymbol{l}^{(\delta)}(\rho_1)| \right) \right] \hat{\boldsymbol{l}}^{(\delta)}(\rho_1) \cdot \hat{\boldsymbol{l}}^{(\delta)} \left(t_{k-1}^{(p)} \right) d\rho_1 \\
&\overset{(2.37)}{\geq} (s - \sigma) h_*(2M_\delta) \left(1 - \frac{2}{N} \right) + \left[h_*(2M_\delta) - \sqrt{\delta} \tilde{D}(2M_\delta) \right] (s - \rho) \left(1 - \frac{2}{N} \right) \\
&\geq (s - \sigma) h_*(2M_\delta) \left(1 - \frac{2}{N} \right),
\end{aligned} \tag{2.134}$$

provided that $\delta \in (0, \delta_0]$ and δ_0 is sufficiently small. We see from (2.134) that (2.39) is satisfied with $r = (1 - 2/N)\, h_*(2M_\delta) \delta^{1-\gamma_A}$. Using Lemma 2.3.2 we estimate

$$\begin{aligned}
\left| \mathbb{E}[J^{(1)}\tilde{\zeta}] \right| &\leq \frac{\tilde{D}(2M_\delta)}{\sqrt{\delta}} \|G\|_1 \mathbb{E}[\tilde{\zeta}] \int_t^u \phi \left(C_A^{(1)} \frac{s - \sigma}{\delta} \right) ds \\
&\leq C_A^{(2)} \|G\|_1 \mathbb{E}[\tilde{\zeta}] \delta^{-1/2} \phi \left(C_A^{(1)} \delta^{-\gamma_A} \right) (u - t) \leq C_A^{(3)} \|G\|_1 \mathbb{E}[\tilde{\zeta}] \delta (u - t)
\end{aligned} \tag{2.135}$$

and $C_A^{(3)}$ exists by virtue of assumption (2.7). On the other hand, the term $J^{(2)}$ defined by (2.133) may be written as $J^{(2)} = J_1^{(2)} + J_2^{(2)}$, where

$$J_1^{(2)} := -\frac{1}{\delta^{3/2}} \sum_{i,j=1}^{d} \int_t^u \partial_j G(\boldsymbol{l}^{(\delta)}(\sigma)) \partial_{y_i} F_{j,\delta} \left(s, \frac{\boldsymbol{L}^{(\delta)}(\sigma, s)}{\delta}, \boldsymbol{l}^{(\delta)}(\sigma) \right) (y_i^{(\delta)}(s) - L_i^{(\delta)}(\sigma, s))\, ds$$

and

$$J_2^{(2)} := -\frac{1}{\delta^{5/2}} \sum_{i,j,k=1}^{d} \int_t^u \int_0^1 \int_0^1 \partial^2_{y_i,y_k} F_{j,\delta}\left(s, \frac{\boldsymbol{R}^{(\delta)}(\theta v, \sigma, s)}{\delta}, \boldsymbol{l}^{(\delta)}(\sigma)\right) v \qquad (2.136)$$
$$\times\, \partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))(y_i^{(\delta)}(s) - L_i^{(\delta)}(\sigma,s))(y_k^{(\delta)}(s) - L_k^{(\delta)}(\sigma,s))\, ds\, dv\, d\theta.$$

The term involving $J_2^{(2)}$ may be handled easily with the help of (2.131) and Lemma 2.3.1. We have then

$$|\mathbb{E}[J_2^{(2)}\tilde{\zeta}]| \le C_A^{(4)} \tilde{D}(2M_\delta)\mathbb{E}[\tilde{\zeta}]\|G\|_1(u-t)\delta^{-5/2}\delta^{3-4\gamma_A-4(\epsilon_2+\epsilon_3)}T^2 \qquad (2.137)$$
$$\le C_A^{(5)}\delta^{1/2-4\gamma_A-4(\epsilon_2+\epsilon_3)}T^2(u-t)\mathbb{E}[\tilde{\zeta}]\|G\|_1.$$

In order to estimate the term corresponding to $J_1^{(2)}$ we write $J_1^{(2)} = J_{1,1}^{(2)} + J_{1,2}^{(2)}$, where

$$J_{1,1}^{(2)} := -\frac{1}{\delta^{3/2}} \sum_{i,j=1}^{d} \int_t^u \int_\sigma^s \partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\partial_{y_i} F_{j,\delta}\left(s, \frac{\boldsymbol{L}^{(\delta)}(\sigma,s)}{\delta}, \boldsymbol{l}^{(\delta)}(\sigma)\right) \qquad (2.138)$$
$$\times\, (s-\rho_1)\frac{d}{d\rho_1}\left[H_0'(|\boldsymbol{l}^{(\delta)}(\rho_1)|)\,\hat{l}_i^{(\delta)}(\rho_1)\right] ds\, d\rho_1$$

and

$$J_{1,2}^{(2)} := -\frac{1}{\delta} \sum_{i,j=1}^{d} \int_t^u \int_\sigma^s \partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\partial_{y_i} F_{j,\delta}\left(s, \frac{\tilde{\boldsymbol{L}}^{(\delta)}(\sigma,s)}{\delta}, \boldsymbol{l}^{(\delta)}(\sigma)\right)$$
$$\times \partial_l H_1\left(\frac{y^{(\delta)}(\rho)}{\delta}, |\boldsymbol{l}^{(\delta)}(\rho)|\right)\hat{l}_i^{(\delta)}(\rho)\, ds\, d\rho,$$

with

$$\frac{d}{d\rho_1}\left[H_0'(|\boldsymbol{l}^{(\delta)}(\rho_1)|)\,\hat{l}_i^{(\delta)}(\rho_1)\right] = H_0''(|\boldsymbol{l}^{(\delta)}(\rho_1)|)\,(\hat{\boldsymbol{l}}^{(\delta)}(\rho_1), \frac{d}{d\rho_1}\boldsymbol{l}^{(\delta)}(\rho_1))_{\mathbb{R}^d}\,\hat{l}_i^{(\delta)}(\rho_1) \qquad (2.139)$$
$$+H_0'(|\boldsymbol{l}^{(\delta)}(\rho_1)|)|\boldsymbol{l}^{(\delta)}(\rho_1)|^{-1}\left[\frac{d}{d\rho_1}l_i^{(\delta)}(\rho_1) - (\hat{\boldsymbol{l}}^{(\delta)}(\rho_1), \frac{d}{d\rho_1}\boldsymbol{l}^{(\delta)}(\rho_1))_{\mathbb{R}^d}\,\hat{l}_i^{(\delta)}(\rho_1)\right].$$

We deal with $J_{1,2}^{(2)}$ first. It may be split as $J_{1,2}^{(2)} = J_{1,2,1}^{(2)} + J_{1,2,2}^{(2)} + J_{1,2,3}^{(2)}$, where

$$J_{1,2,1}^{(2)} := -\frac{1}{\delta} \sum_{i,j=1}^{d} \int_t^u (s-\sigma)\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\partial_{y_i} F_{j,\delta}\left(s, \frac{\boldsymbol{L}^{(\delta)}(\sigma,s)}{\delta}, \boldsymbol{l}^{(\delta)}(\sigma)\right) \qquad (2.140)$$
$$\times\, \partial_l H_1\left(\frac{\boldsymbol{L}^{(\delta)}(\sigma,\rho)}{\delta}, |\boldsymbol{l}^{(\delta)}(\sigma)|\right)\hat{l}_i^{(\delta)}(\sigma)\, ds$$

$$J_{1,2,2}^{(2)} := -\frac{1}{\delta^2} \sum_{i,j=1}^{d} \int_t^u \int_\sigma^s \int_0^1 \partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\partial_{y_i} F_{j,\delta}\left(s, \frac{\boldsymbol{L}^{(\delta)}(\sigma,s)}{\delta}, \boldsymbol{l}^{(\delta)}(\sigma)\right)$$
$$\times\, (\partial_{y_i}\partial_l H_1)\left(\frac{\boldsymbol{R}^{(\delta)}(v,\sigma,\rho)}{\delta}, |\boldsymbol{l}^{(\delta)}(\rho)|\right)(y_i^{(\delta)}(\rho) - L_i^{(\delta)}(\sigma,\rho))\,\hat{l}_i^{(\delta)}(\rho)\, ds\, d\rho\, dv$$

and

$$J_{1,2,3}^{(2)} := -\frac{1}{\delta}\sum_{i,j=1}^{d}\int_t^u\int_\sigma^s\int_\sigma^\rho \partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\partial_{y_i}F_{j,\delta}\left(s, \frac{\boldsymbol{L}^{(\delta)}(\sigma,s)}{\delta}, \boldsymbol{l}^{(\delta)}(\sigma)\right)$$
$$\times \frac{d}{d\rho_1}\left[\partial_l H_1\left(\frac{\boldsymbol{L}^{(\delta)}(\sigma,s)}{\delta}, |\boldsymbol{l}^{(\delta)}(\rho_1)|\right)\hat{l}_i^{(\delta)}(\rho_1)\right] ds\, d\rho\, d\rho_1.$$

By virtue of (2.131), definition (2.35) and Lemma 2.3.1 we obtain easily that

$$|\mathbb{E}[J_{1,2,2}^{(2)}\tilde{\zeta}]| \le C_A^{(6)}\delta^{1/2-(3\gamma_A+2\epsilon_2+2\epsilon_3)}\|G\|_1 T(u-t)\mathbb{E}\tilde{\zeta}. \tag{2.141}$$

The same argument and equality (2.139) also allow us to estimate $|\mathbb{E}[J_{1,2,3}^{(2)}\zeta]|$ by the right hand side of (2.141).

Using Lemma 2.3.1 and the definition (2.35) we conclude that there exists a constant $C_A^{(7)} > 0$ independent of δ such that

$$\left|\partial_{y_i}F_{j,\delta}\left(s, \frac{\boldsymbol{L}^{(\delta)}(\sigma,s)}{\delta}, \boldsymbol{l}^{(\delta)}(\sigma)\right) - \Theta\left(t_k^{(p)}, \boldsymbol{L}^{(\delta)}(\sigma,s), \boldsymbol{l}^{(\delta)}(\sigma)\right)\partial^2_{y_i,y_j}H_1\left(\frac{\boldsymbol{L}^{(\delta)}(\sigma,s)}{\delta}, |\boldsymbol{l}^{(\delta)}(\sigma)|\right)\right|$$
$$\le C_A^{(7)}\delta^{1-2(\epsilon_2+\epsilon_3)}T, \quad i,j = 1,\ldots,d.$$

Therefore, we can write

$$\Bigg|\mathbb{E}[J_{1,2,1}^{(2)}\tilde{\zeta}] + \frac{1}{\delta}\sum_{i,j=1}^{d}\int_t^u\int_\sigma^s \mathbb{E}\left[\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\Theta\left(t_k^{(p)}, \boldsymbol{L}^{(\delta)}(\sigma,s), \boldsymbol{l}^{(\delta)}(\sigma)\right)\right. \tag{2.142}$$
$$\left.\times\, \partial^2_{y_i,y_j}H_1\left(\frac{\boldsymbol{L}^{(\delta)}(\sigma,s)}{\delta}, |\boldsymbol{l}^{(\delta)}(\sigma)|\right)\partial_l H_1\left(\frac{\boldsymbol{L}^{(\delta)}(\sigma,\rho)}{\delta}, |\boldsymbol{l}^{(\delta)}(\sigma)|\right)\hat{l}_i^{(\delta)}(\sigma)\,\tilde{\zeta}\right] ds\, d\rho\Bigg|$$
$$\le C_A^{(8)}\delta^{1-\gamma_A-2(\epsilon_2+\epsilon_3)}(u-t)\|G\|_1 T\mathbb{E}\tilde{\zeta}.$$

With our choice of the exponents we have $\delta < (2q)^{[1-\gamma_A-2(\epsilon_2+\epsilon_3)]^{-1}}$ for all $\delta \in [0,\delta_0)$ where $\delta_0 > 0$ is sufficiently small. We apply now part (ii) of Lemma 2.3.2 with

$$Z = \partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\Theta\left(t_k^{(p)}, \boldsymbol{L}^{(\delta)}(\sigma,s), \boldsymbol{l}^{(\delta)}(\sigma)\right)\hat{l}_i^{(\delta)}(\sigma)\,\tilde{\zeta},$$
$$\tilde{X}_1(x,k) := \partial^2_{x_i,x_j}H_1(\mathbf{x},k), \quad \tilde{X}_2(x) := \partial_k H_1(\mathbf{x},k),$$
$$g_1 := \left(\frac{\boldsymbol{L}^{(\delta)}(\sigma,s)}{\delta}, |\boldsymbol{l}^{(\delta)}(\sigma)|\right), \quad g_2 := \left(\frac{\boldsymbol{L}^{(\delta)}(\sigma,\rho)}{\delta}, |\boldsymbol{l}^{(\delta)}(\sigma)|\right),$$
$$r = C_A^{(9)}(\rho-\sigma), \quad r_1 = C_A^{(9)}(s-\rho).$$

We conclude that

$$\Bigg|\mathbb{E}\left[J_{1,2,1}^{(2)}\zeta\right] + \frac{1}{\delta}\sum_{i,j=1}^{d}\int_t^u\int_\sigma^s \mathbb{E}\left[\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\Theta\left(t_k^{(p)}, \boldsymbol{L}^{(\delta)}(\sigma,s), \boldsymbol{l}^{(\delta)}(\sigma)\right)\right. \tag{2.143}$$
$$\left.\times\partial^2_{y_i,y_j}R_1\left(\frac{\boldsymbol{L}^{(\delta)}(\sigma,s)-\boldsymbol{L}^{(\delta)}(\sigma,\rho)}{\delta},\ |\boldsymbol{l}^{(\delta)}(\sigma)|\right)\hat{l}_i^{(\delta)}(\sigma)\tilde{\zeta}\right] ds\, d\rho\Bigg|$$
$$\le C_A^{(8)}\delta^{1-\gamma_A-2(\epsilon_2+\epsilon_3)}(u-t)\|G\|_1 T\mathbb{E}\tilde{\zeta}$$
$$+\frac{C_A^{(10)}}{\delta}\|G\|_1\mathbb{E}[\tilde{\zeta}]\int_t^u\int_\sigma^s \phi^{1/2}\left(\frac{C_A^{(9)}(\rho-\sigma)}{2\delta}\right)\phi^{1/2}\left(\frac{C_A^{(9)}(s-\rho)}{2\delta}\right) ds\, d\rho,$$

where

$$R_1(\mathbf{y},k) := \mathbb{E}[H_1(\mathbf{y},k)\partial_k H_1(\mathbf{0},k)], \quad (\mathbf{y},k)\in\mathbb{R}^d\times[0,+\infty). \tag{2.144}$$

We can use assumption (2.7) to estimate the second term on the right hand side of (2.143) e.g. by $C_A^{(11)}\delta(u-t)\|G\|_1\mathbb{E}\tilde{\zeta}$. The second term appearing on the left hand side of (2.143) equals to

$$\sum_{j=1}^d \int_t^u \mathbb{E}\left\{\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\Theta\left(t_k^{(p)},\boldsymbol{L}^{(\delta)}(\sigma,s),\boldsymbol{l}^{(\delta)}(\sigma)\right)\right. \tag{2.145}$$

$$\left.\times\frac{1}{H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)}\left[-\int_\sigma^s \frac{d}{d\rho}\partial_{y_j}R_1\left(\frac{s-\rho}{\delta}\,H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)\,\hat{\boldsymbol{l}}^{(\delta)}(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\right)d\rho\right]\tilde{\zeta}\right\}ds$$

and integrating over $d\rho$ we obtain that it equals

$$-\sum_{j=1}^d \int_t^u \mathbb{E}\left\{\frac{\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))}{H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)}\Theta\left(t_k^{(p)},\boldsymbol{L}^{(\delta)}(\sigma,s),\boldsymbol{l}^{(\delta)}(\sigma)\right)\partial_{y_j}R_1\left(\mathbf{0},|\boldsymbol{l}^{(\delta)}(\sigma)|\right)\tilde{\zeta}\right\}ds \tag{2.146}$$

$$+\sum_{j=1}^d \int_t^u \mathbb{E}\left\{\frac{\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))}{H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)}\Theta\left(t_k^{(p)},\boldsymbol{L}^{(\delta)}(\sigma,s),\boldsymbol{l}^{(\delta)}(\sigma)\right)\partial_{y_j}R_1\left(\delta^{-\gamma_A}\,\hat{\boldsymbol{l}}^{(\delta)}(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\right)\tilde{\zeta}\right\}ds.$$

By virtue of (2.8) the second term appearing in (2.146) is bounded e.g. by $C_A^{(12)}\delta(u-t)\|G\|_1\mathbb{E}\tilde{\zeta}$ for some constant $C_A^{(12)}>0$, thus we have shown that

$$\left|\mathbb{E}[J_{1,2,1}^{(2)}\zeta]-\sum_{j=1}^d \int_t^u \mathbb{E}\left\{\frac{\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))}{H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)}\Theta\left(t_k^{(p)},\boldsymbol{L}^{(\delta)}(\sigma,s),\boldsymbol{l}^{(\delta)}(\sigma)\right)\partial_{y_j}R_1\left(\mathbf{0},|\boldsymbol{l}^{(\delta)}(\sigma)|\right)\tilde{\zeta}\right\}ds\right| \tag{2.147}$$

$$\le C_A^{(13)}\delta^{1-\gamma_A-2(\epsilon_2+\epsilon_3)}(u-t)\|G\|_1 T\mathbb{E}\tilde{\zeta}.$$

Let us consider the term corresponding to $J_{1,1}^{(2)}$, cf. (2.138). Note that according to (2.139) and (2.36) we have $J_{1,1}^{(2)}=J_{1,1,1}^{(2)}+J_{1,1,2}^{(2)}$, where

$$J_{1,1,1}^{(2)} := -\frac{1}{\delta^2}\sum_{i,j=1}^d \int_t^u\int_\sigma^s \partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\partial_{y_i}F_{j,\delta}\left(s,\frac{\boldsymbol{L}^{(\delta)}(\sigma,s)}{\delta},\boldsymbol{l}^{(\delta)}(\sigma)\right)$$

$$\times(s-\rho_1)\Gamma_i\left(\rho_1,\frac{y^{(\delta)}(\rho_1)}{\delta},\boldsymbol{l}^{(\delta)}(\sigma)\right)ds\,d\rho_1,$$

with

$$\Gamma_i(\rho,\mathbf{y},\mathbf{l}) := |\mathbf{l}|^{-1}H_0'(|\mathbf{l}|)\left[\left(\hat{\mathbf{l}},F_\delta(\rho,\mathbf{y},\mathbf{l})\right)_{\mathbb{R}^d}l_i - F_{i,\delta}(\rho,\mathbf{y},\mathbf{l})\right]-H_0''(|\mathbf{l}|)\left(\hat{\mathbf{l}},F_\delta(\rho,\mathbf{y},\mathbf{l})\right)_{\mathbb{R}^d}\hat{l}_i,$$

while

$$J_{1,1,2}^{(2)} := -\frac{1}{\delta^2}\sum_{i,j=1}^d \int_t^u\int_\sigma^s\int_\sigma^{\rho_1} \partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\partial_{y_i}F_{j,\delta}\left(s,\frac{\boldsymbol{L}^{(\delta)}(\sigma,s)}{\delta},\boldsymbol{l}^{(\delta)}(\sigma)\right)$$

$$\times(s-\rho_1)\frac{d}{d\rho_2}\Gamma_i\left(\rho_1,\frac{y^{(\delta)}(\rho_1)}{\delta},\boldsymbol{l}^{(\delta)}(\rho_2)\right)ds\,d\rho_1\,d\rho_2. \tag{2.148}$$

Note that $|\frac{d}{d\rho_2}\Gamma_i| \le C_A^{(14)}\delta^{-1/2}$ for some constant $C_A^{(14)} > 0$. A straightforward computation, using (2.130) and Lemma 2.3.1, shows that $|\mathbb{E}[J_{1,1,2}^{(2)}\zeta]| \le C_A^{(15)}\delta^{1/2-(3\gamma_A+2\epsilon_2+2\epsilon_3)}(u-t)\|G\|_1 T\mathbb{E}[\tilde{\zeta}]$. An application of (2.131), in the same fashion as it was done in the calculations concerning the terms $\mathbb{E}[J_{1,2,2}^{(2)}\zeta]$ and $\mathbb{E}[J_{1,2,3}^{(2)}\zeta]$, yields that

$$\left|\mathbb{E}[J_{1,1,1}^{(2)}\zeta] + \frac{1}{\delta^2}\sum_{i,j=1}^{d}\int_t^u\int_\sigma^s (s-\rho_1)\mathbb{E}\left[\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\partial_{y_i}F_{j,\delta}\left(s, \frac{\boldsymbol{L}^{(\delta)}(\sigma,s)}{\delta}, \boldsymbol{l}^{(\delta)}(\sigma)\right)\right.\right. \quad (2.149)$$
$$\left.\left.\times\ \Gamma_i\left(\rho_1, \frac{\boldsymbol{L}^{(\delta)}(\sigma,\rho_1)}{\delta}, \boldsymbol{l}^{(\delta)}(\sigma)\right)\tilde{\zeta}\right] ds\, d\rho_1\right| \le C_A^{(16)}\delta^{1/2-(4\gamma_A+2\epsilon_2+2\epsilon_3)}(u-t)\|G\|_1 T\mathbb{E}[\tilde{\zeta}].$$

For $j = 1, \ldots, d$ we let

$$V_j(\mathbf{y},\mathbf{y}',\mathbf{l}) := \sum_{i,k=1}^{d}\left(H_0''(|\mathbf{l}|) - H_0'(|\mathbf{l}|)\right)\partial^3_{y_i,y_j,y_k}R(\mathbf{y}-\mathbf{y}',|\mathbf{l}|)\hat{l}_i\hat{l}_k$$
$$+\sum_{i=1}^{d} H_0'(|\mathbf{l}|)|\mathbf{l}|^{-1}\partial^3_{y_i,y_i,y_j}R(\mathbf{y}-\mathbf{y}',|\mathbf{l}|),$$

and also

$$\Lambda(t,\mathbf{y},\mathbf{y}',\mathbf{l};\pi) := \Theta(t,\mathbf{y},\mathbf{l};\pi)\Theta(t,\mathbf{y}',\mathbf{l};\pi), \quad t \ge 0,\ \mathbf{y},\mathbf{y}' \in \mathbb{R}^d,\ \mathbf{l} \in \mathbb{R}^d_*,\ \pi \in \mathcal{C}, \qquad (2.150)$$

$P := \left(\boldsymbol{L}^{(\delta)}(\sigma,s), \boldsymbol{L}^{(\delta)}(\sigma,\rho_1), \boldsymbol{l}^{(\delta)}(\sigma)\right)$, $P_\delta := \left(\delta^{-1}\boldsymbol{L}^{(\delta)}(\sigma,s), \delta^{-1}\boldsymbol{L}^{(\delta)}(\sigma,\rho_1), \boldsymbol{l}^{(\delta)}(\sigma)\right)$ and

$$\overline{\Theta}(s) := \Theta(s, y^{(\delta)}(s), \boldsymbol{l}^{(\delta)}(s); y^{(\delta)}(\cdot), \boldsymbol{l}^{(\delta)}(\cdot)).$$

Applying Lemma 2.3.1 and part ii) of Lemma 2.3.2, as in (2.142) and (2.143), we conclude that the difference between the second term on the left hand side of (2.149) and

$$\frac{1}{\delta^2}\sum_{j=1}^{d}\int_t^u\int_\sigma^s (s-\rho_1)\mathbb{E}\left[\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\Lambda(\sigma,P)V_j(P_\delta)\ \tilde{\zeta}\right] ds\, d\rho_1, \qquad (2.151)$$

can be estimated by $C_A^{(17)}\delta^{\gamma_A^{(1)}}(u-t)\|G\|_1\mathbb{E}[\tilde{\zeta}]$ for some $\gamma_A^{(1)} > 0$. Using the fact that

$$|\boldsymbol{l}^{(\delta)}(\rho) - \boldsymbol{l}^{(\delta)}(\sigma)| \le C_A^{(22)}\delta^{1/2-\gamma_A}, \quad \rho \in [\sigma, s], \qquad (2.152)$$

estimate (2.131) and Lemma 2.3.1 we can argue that

$$\left|\Lambda(\sigma,P) - \overline{\Theta}^2(s)\right| \le C_A^{(18)}(\delta^{1/2-\gamma_A-\epsilon_1} + \delta^{1/2-2(\gamma_A+\epsilon_2+\epsilon_3)}T).$$

We conclude therefore that the magnitude of the difference between the expression in (2.151) and

$$\frac{1}{\delta^2}\sum_{j=1}^{d}\int_t^u \mathbb{E}\left[\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\overline{\Theta}^2(s)\left(\int_\sigma^s (s-\rho_1)V_j(P_\delta)\, d\rho_1\right)\tilde{\zeta}\right] ds, \qquad (2.153)$$

can be estimated by $C_A^{(19)}\delta^{\gamma_A^{(2)}}(u-t)\|G\|_1 T\mathbb{E}[\tilde{\zeta}]$ for some $\gamma_A^{(2)} > 0$. Using shorthand notation $Q(\sigma) := H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)\,\hat{\boldsymbol{l}}^{(\delta)}(\sigma)$ we can write the integral from σ to s appearing above as being equal to

$$\frac{1}{\delta^2}\int\limits_{s-\delta^{1-\gamma_A}}^{s}(s-\rho_1)\left[\sum_{i,k=1}^{d}\Big(H_0''(|\boldsymbol{l}^{(\delta)}(\sigma)|)-H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)\Big)\,\partial^3_{y_i,y_j,y_k}R\left(\frac{s-\rho_1}{\delta}\,Q(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\right)\right.$$

$$\left.\times\hat{l}_i^{(\delta)}(\sigma)\hat{l}_k^{(\delta)}(\sigma)+H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)|\boldsymbol{l}^{(\delta)}(\sigma)|^{-1}\sum_{i=1}^{d}\partial^3_{y_i,y_i,y_j}R\left(\frac{s-\rho_1}{\delta}\,Q(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\right)\right]d\rho_1,$$

which upon the change of variables $\rho_1 := (s-\rho_1)/\delta$ is equal to

$$\int\limits_{0}^{\delta^{-\gamma_A}}\rho_1\left[\sum_{i,k=1}^{d}\Big(H_0''(|\boldsymbol{l}^{(\delta)}(\sigma)|)-H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)\Big)\,\partial^3_{y_i,y_j,y_k}R\Big(\rho_1\,Q(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\Big)\,\hat{l}_i^{(\delta)}(\sigma)\hat{l}_k^{(\delta)}(\sigma)\right.$$

$$\left.+H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)|\boldsymbol{l}^{(\delta)}(\sigma)|^{-1}\sum_{i=1}^{d}\partial^3_{y_i,y_i,y_j}R\Big(\rho_1\,Q(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\Big)\right]d\rho_1. \tag{2.154}$$

Using the fact that

$$\sum_{k=1}^{d}\partial^3_{y_i,y_j,y_k}R\Big(\rho_1\,Q(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\Big)\,\hat{l}_k^{(\delta)}(\sigma)=\frac{1}{H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)}\frac{d}{d\rho_1}\Big[\partial^2_{y_i,y_j}R\Big(\rho_1\,Q(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\Big)\Big]$$

we obtain, upon integrating by parts in the first term on the right hand side of (2.154), that this expression equals

$$H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)^{-1}\Big(H_0''(|\boldsymbol{l}^{(\delta)}(\sigma)|)-H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)\Big)\sum_{i=1}^{d}\left[\delta^{-\gamma_A}\partial^2_{y_i,y_j}R\Big(\delta^{-\gamma_A}\,Q(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\Big)\,\hat{l}_i^{(\delta)}(\sigma)\right.$$

$$\left.-\int\limits_{0}^{\delta^{-\gamma_A}}\partial^2_{y_i,y_j}R\Big(\rho_1\,Q(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\Big)\,\hat{l}_i^{(\delta)}(\sigma)\,d\rho_1\right] \tag{2.155}$$

$$+H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)|\boldsymbol{l}^{(\delta)}(\sigma)|^{-1}\int\limits_{0}^{\delta^{-\gamma_A}}\rho_1\partial^3_{y_i,y_i,y_j}R\Big(\rho_1\,Q(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\Big)\,d\rho_1. \tag{2.156}$$

Note that $\nabla R(\mathbf{0}) = \mathbf{0}$ and

$$\sum_{i=1}^{d}\partial^2_{y_i,y_j}R\Big(\rho_1\,Q(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\Big)\,\hat{l}_i^{(\delta)}(\sigma)=\frac{1}{H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)}\frac{d}{d\rho_1}\partial_{y_j}R\Big(\rho_1\,Q(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\Big).$$

We obtain therefore that the expression in (2.155) equals

$$H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)^{-1}\left(H_0''(|\boldsymbol{l}^{(\delta)}(\sigma)|)-H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)\right)\left[\sum_{i=1}^{d}\delta^{-\gamma_A}\partial^2_{y_i,y_j}R\left(\delta^{-\gamma_A}Q(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\right)\hat{l}_i^{(\delta)}(\sigma)\right.$$

$$\left.-H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)^{-1}\partial_{y_j}R\left(\delta^{-\gamma_A}Q(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\right)\right] \tag{2.157}$$

$$+H_0'(|\boldsymbol{l}^{(\delta)}(\sigma)|)|\boldsymbol{l}^{(\delta)}(\sigma)|^{-1}\sum_{i=1}^{d}\int_0^{\delta^{-\gamma_A}}\rho_1\partial^3_{y_i,y_i,y_j}R\left(\rho_1 Q(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|\right)d\rho_1.$$

Recalling assumption (2.8) we conclude that the expressions corresponding to the first two terms appearing in (2.157) are of order of magnitude $O(\delta^{\gamma_A^{(3)}})$ for some $\gamma_A^{(3)}>0$. Summarizing work done in this section, we have shown that

$$\left|\mathbb{E}\left\{\left[I^{(1)}-\sum_{j=1}^{d}\int_t^u C_j(\boldsymbol{l}^{(\delta)}(\sigma))\overline{\Theta}^2(s)\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\,ds\right]\tilde{\zeta}\right\}\right|\le C_A^{(20)}\delta^{\gamma_A^{(4)}}(u-t)\|G\|_1T^2\mathbb{E}\tilde{\zeta} \tag{2.158}$$

for some constants $C_A^{(20)},\gamma_A^{(4)}>0$ and (cf. (2.144))

$$C_j(\mathbf{l}):=E_j(\hat{\mathbf{l}},|\mathbf{l}|)+\frac{\partial_{y_j}R_1(\mathbf{0},|\mathbf{l}|)}{H_0'(|\mathbf{l}|)},$$

$$E_j(\hat{\mathbf{l}},k):=-\frac{H_0'(k)}{k}\sum_{i=1}^{d}\int_0^{+\infty}\rho_1\partial^3_{y_i,y_i,y_j}R\left(\rho_1 H_0'(k)\hat{\mathbf{l}},k\right)d\rho_1,\quad j=1,\ldots,d.$$

The terms $\mathbb{E}[I^{(2)}\tilde{\zeta}]$ and $\mathbb{E}[I^{(3)}\tilde{\zeta}]$

The calculations concerning these terms essentially follow the respective steps performed in the previous section so we only highlight their main points. First, we note that the difference between $\mathbb{E}[I^{(2)}\tilde{\zeta}]$ and

$$\frac{1}{\delta}\sum_{i,j=1}^{d}\int_t^u\int_\sigma^s\mathbb{E}\left[\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\partial_{\ell_i}F_{j,\delta}\left(s,\frac{y^{(\delta)}(s)}{\delta},\boldsymbol{l}^{(\delta)}(\sigma)\right)F_{i,\delta}\left(\rho,\frac{y^{(\delta)}(\rho)}{\delta},\boldsymbol{l}^{(\delta)}(\sigma)\right)\tilde{\zeta}\right]ds\,d\rho \tag{2.159}$$

is less than, or equal to $C_A^{(21)}\delta^{\gamma_A^{(5)}}(u-t)\|G\|_1\mathbb{E}[\tilde{\zeta}]$, cf. (2.152). Next we note that (2.159) equals

$$\frac{1}{\delta}\sum_{i,j=1}^{d}\int_t^u\int_\sigma^s \mathbb{E}\left[\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\partial_{\ell_i}F_{j,\delta}\left(s,\frac{\boldsymbol{L}^{(\delta)}(\sigma,s)}{\delta},\boldsymbol{l}^{(\delta)}(\sigma)\right)F_{i,\delta}\left(\rho,\frac{\boldsymbol{L}^{(\delta)}(\sigma,\rho)}{\delta},\boldsymbol{l}^{(\delta)}(\sigma)\right)\tilde{\zeta}\right]ds\,d\rho$$

$$+\frac{1}{\delta^2}\sum_{i,j,k=1}^{d}\int_t^u\int_\sigma^s\int_0^1 \mathbb{E}\left[\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\partial_{\ell_i}\partial_{y_k}F_{j,\delta}\left(s,\frac{\boldsymbol{R}^{(\delta)}(v,\sigma,s)}{\delta},\boldsymbol{l}^{(\delta)}(\sigma)\right)\right. \tag{2.160}$$

$$\left.\times F_{i,\delta}\left(\rho,\frac{\boldsymbol{L}^{(\delta)}(\sigma,\rho)}{\delta},\boldsymbol{l}^{(\delta)}(\sigma)\right)(y_k^{(\delta)}(s)-L_k^{(\delta)}(\sigma,s))\tilde{\zeta}\right]ds\,d\rho\,dv$$

$$+\frac{1}{\delta^2}\sum_{i,j,k=1}^{d}\int_t^u\int_\sigma^s\int_0^1 \mathbb{E}\left[\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\partial_{\ell_i}F_{j,\delta}\left(s,\frac{y^{(\delta)}(s)}{\delta},\boldsymbol{l}^{(\delta)}(\sigma)\right)\right.$$

$$\left.\times\,\partial_{y_k}F_{i,\delta}\left(\rho,\frac{\boldsymbol{R}^{(\delta)}(v,\sigma,\rho)}{\delta},\boldsymbol{l}^{(\delta)}(\sigma)\right)(y_k^{(\delta)}(\rho)-L_k^{(\delta)}(\sigma,\rho))\tilde{\zeta}\right]ds\,d\rho\,dv.$$

A straightforward argument using Lemma 2.3.1 and (2.131) shows that both the second and third terms of (2.160) can be estimated by $C_A^{(23)}\delta^{1/2-(3\gamma_A+2\epsilon_2+2\epsilon_3)}(u-t)\|G\|_1T^2\mathbb{E}[\tilde{\zeta}]$. The first term, on the other hand, can be handled with the help of part ii) of Lemma 2.3.2 in the same fashion as we have dealt with the term $J_{1,2,1}^{(2)}$, given by (2.140) of Section 2.6, and we obtain that

$$\left|\mathbb{E}\left\{\left[I^{(2)}-\sum_{j=1}^{d}\int_t^u\left(D_j(|\boldsymbol{l}^{(\delta)}(\sigma)|)\overline{\Theta}^2(s)+J_j(s;y^{(\delta)}(\cdot),\boldsymbol{l}^{(\delta)}(\cdot))\overline{\Theta}(s)\right)\partial_j G(\boldsymbol{l}^{(\delta)}(\sigma))\,ds\right]\tilde{\zeta}\right\}\right| \tag{2.161}$$

$$\le C_A^{(24)}\delta^{\gamma_A^{(6)}}(u-t)\|G\|_1T\mathbb{E}[\tilde{\zeta}].$$

Here

$$D_j(l):=\frac{\partial_{y_j}R_2(\mathbf{0},l)}{H_0'(l)},\qquad R_2(\mathbf{y},l):=\mathbb{E}[\partial_l H_1(\mathbf{y},l)H_1(\mathbf{0},l)], \tag{2.162}$$

$$J_j(s;y^{(\delta)}(\cdot),\boldsymbol{l}^{(\delta)}(\cdot)):=-\sum_{i=1}^{d}\overline{\Theta}_i(s)D_{i,j}(\hat{\boldsymbol{l}}^{(\delta)}(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|),$$

$$\overline{\Theta}_i(s):=\partial_{l_i}\Theta(s,y^{(\delta)}(s),\boldsymbol{l}^{(\delta)}(s);y^{(\delta)}(\cdot),\boldsymbol{l}^{(\delta)}(\cdot)).$$

Finally, concerning the limit of $\mathbb{E}[I^{(3)}\tilde{\zeta}]$, another application of (2.131) yields

$$\left|\mathbb{E}[I^{(3)}\tilde{\zeta}]-\mathcal{I}\right|\le C_A^{(25)}\delta^{\gamma_A^{(7)}}(u-t)\|G\|_1\mathbb{E}[\tilde{\zeta}], \tag{2.163}$$

where

$$\mathcal{I}:=\frac{1}{\delta}\int_t^u\int_\sigma^s\mathbb{E}\left\{\partial_{i,j}^2 G(\boldsymbol{l}^{(\delta)}(\sigma))F_{j,\delta}\left(s,\frac{\boldsymbol{L}^{(\delta)}(\sigma,s)}{\delta},\boldsymbol{l}^{(\delta)}(\sigma)\right)F_{i,\delta}\left(\rho,\frac{\boldsymbol{L}^{(\delta)}(\sigma,\rho)}{\delta},\boldsymbol{l}^{(\delta)}(\sigma)\right)\tilde{\zeta}\right\}ds\,d\rho.$$

Then, we can use part ii) of Lemma 2.3.2 in order to obtain

$$\left|\mathcal{I}-\sum_{i,j=1}^{d}\int_t^u D_{i,j}(\hat{\boldsymbol{l}}^{(\delta)}(\sigma),|\boldsymbol{l}^{(\delta)}(\sigma)|)\overline{\Theta}^2(s)\partial_{i,j}^2G(\boldsymbol{l}^{(\delta)}(\sigma))\,ds\right|\le C_A^{(26)}\delta^{\gamma_A^{(8)}}(u-t)\|G\|_2T\mathbb{E}[\tilde{\zeta}]. \tag{2.164}$$

Next we replace the argument σ, in formulas (2.158), (2.161) and (2.163), by s. This can be done thanks to estimate (2.152) and the assumption on the regularity of the random field $H_1(\cdot,\cdot)$. In order to make this approximation work we will be forced to use the third derivative of $G(\cdot)$.

Finally (cf. (2.144), (2.162)) note that

$$\nabla_{\mathbf{y}} R_1(\mathbf{0},l) + \nabla_{\mathbf{y}} R_2(\mathbf{0},l) = \nabla_{\mathbf{y}\,|\,\mathbf{y}=\mathbf{0}} \mathbb{E}[\partial_l H_1(\mathbf{y},l) H_1(\mathbf{y},l)] = \mathbf{0}.$$

Hence we conclude that the assertion of the lemma holds for any function $G \in C^3(\mathbb{R}^d_*)$ satisfying $\|G\|_3 < +\infty$. Generalization to an arbitrary $G \in C_b^{1,1,3}([0,+\infty) \times \mathbb{R}^{2d}_*)$ is fairly standard. Let r_0 be any integer and consider $s_k := t + kr_0^{-1}(u-t)$, $k = 0, \ldots, r_0$. Then

$$\mathbb{E}\left\{[G(u, y^{(\delta)}(u), \boldsymbol{l}^{(\delta)}(u)) - G(t, y^{(\delta)}(t), \boldsymbol{l}^{(\delta)}(t))]\tilde{\zeta}\right\} \tag{2.165}$$

$$= \sum_{k=0}^{r_0-1} \mathbb{E}\left\{[G(s_{k+1}, y^{(\delta)}(s_{k+1}), \boldsymbol{l}^{(\delta)}(s_{k+1})) - G(s_k, y^{(\delta)}(s_k), \boldsymbol{l}^{(\delta)}(s_k))]\tilde{\zeta}\right\}.$$

$$= \sum_{k=0}^{r_0-1} \mathbb{E}\left\{[G(s_k, y^{(\delta)}(s_k), \boldsymbol{l}^{(\delta)}(s_{k+1})) - G(s_k, y^{(\delta)}(s_k), \boldsymbol{l}^{(\delta)}(s_k))]\tilde{\zeta}\right\}$$

$$+ \sum_{k=0}^{r_0-1} \mathbb{E}\left\{[G(s_{k+1}, y^{(\delta)}(s_{k+1}), \boldsymbol{l}^{(\delta)}(s_k)) - G(s_k, y^{(\delta)}(s_k), \boldsymbol{l}^{(\delta)}(s_k))]\tilde{\zeta}\right\}$$

Using the already proven part of the lemma we obtain

$$\left|\sum_{k=0}^{r_0-1} \mathbb{E}\left\{[\widehat{N}_{s_{k+1}}(G(s_k, y^{(\delta)}(s_k), \cdot)) - \widehat{N}_{s_k}(G(s_k, y^{(\delta)}(s_k), \cdot))]\tilde{\zeta}\right\}\right| \tag{2.166}$$

$$\leq C_A^{(27)} \delta^{\gamma_A^{(9)}} (u-t) \|G\|_{1,1,3} T^2 \mathbb{E}\tilde{\zeta}.$$

On the other hand, the second term on the right hand side of (2.165) equals

$$\sum_{k=0}^{r_0-1} \mathbb{E}\left\{\int\limits_{s_k}^{s_{k+1}} \left\{\partial_\rho + \left[H_0'(|\boldsymbol{l}^{(\delta)}(\rho)|) + \sqrt{\delta}\partial_l H_1\left(\frac{y^{(\delta)}(\rho)}{\delta}, |\boldsymbol{l}^{(\delta)}(\rho)|\right)\right] \hat{\boldsymbol{l}}^{(\delta)}(\rho) \cdot \nabla_{\mathbf{y}}\right\}\right. \tag{2.167}$$

$$\left.\times\, G(\rho, y^{(\delta)}(\rho), \boldsymbol{l}^{(\delta)}(s_k))\tilde{\zeta}\, d\rho\right\}$$

The conclusion of the lemma for an arbitrary function $G \in C_b^{1,1,3}([0,+\infty) \times \mathbb{R}^{2d}_*)$ is an easy consequence of (2.165)–(2.167) upon passing to the limit with $r_0 \to +\infty$. □

Chapter 3

The parabolic regime

The geometric optics regime described in the previous chapter fails when inhomegeneities of the random medium vary on a scale comparable to the wave length. Then the ray theory no longer applies and kinetic description in terms of the Fokker-Planck equation is not applicable. Wave direction undergoes not a continuous diffusive process in the angular variable but rather becomes a jump process. This is the regime of radiative transfer when angularly resolved energy obeys an equation of the form

$$\frac{\partial W(t,x,k)}{\partial t} + c\hat{k}\cdot\nabla_x W(t,x,k) = \int \sigma(k,p)W(x,p)dp - \Sigma(x,k)W(x,k).$$

The differential scattering cross-section $\sigma(k,p)$ and the total scattering cross-section $\Sigma(x,k)$ are determined in terms of the properties of the random medium. The derivation of the macroscopic transport equations from the microscopic wave equations is a very difficult and to a great extent open problem. We will consider in this section the simplest case when fluctuations are "time-dependent" – this dramatically reduces the complexity of rhe problem.

3.1 Derivation of the parabolic wave equation

Let us consider the scalar wave equation for the pressure field $p(z,x,t)$:

$$\frac{1}{c^2(z,x)}\frac{\partial^2 p}{\partial t^2} - \Delta p = 0. \tag{3.1}$$

Here $c(z,x)$ is the local wave speed that we will assume to be random, and the Laplacian operator includes both direction of propagation, z, and the transverse variable $x \in \mathbb{R}^d$. In the physical setting, we have $d = 2$. We consider dimensions $d \geq 1$ to stress that the analysis of the problem is independent of the number of transverse dimensions. If we assume that at time $t = 0$, the wave field has a "beam-like" structure in the z direction, and if back-scattering may be neglected, we can replace the wave equation by its parabolic (also known as paraxial) approximation [59]. More precisely, the pressure p may be approximated as

$$p(z,x,t) \approx \int_{\mathbb{R}} e^{i(-c_0\kappa t+\kappa z)}\psi(z,x,\kappa)c_0 d\kappa, \tag{3.2}$$

where the function ψ satisfies the Schrödinger equation

$$\begin{aligned} &2i\kappa\frac{\partial\psi}{\partial z}(z,x,\kappa) + \Delta_x\psi(z,x,\kappa) + \kappa^2(n^2(z,x)-1)\psi(z,x,\kappa) = 0,\\ &\psi(z=0,x,\kappa) = \psi_0(x,\kappa), \end{aligned} \tag{3.3}$$

with Δ_x the transverse Laplacian in the variable x. The index of refraction $n(z,x) = c_0/c(z,x)$, and c_0 in (3.2) is a reference speed.

A formal justification for the above approximation goes as follows. We start with the reduced wave equation

$$\Delta \hat{p} + \kappa^2 n^2(z,x)\hat{p} = 0, \tag{3.4}$$

and look for solutions of (3.4) in the form $\hat{p}(z,x) = e^{i\kappa z}\psi(z,x)$. We obtain that

$$\frac{\partial^2 \psi}{\partial z^2} + 2i\kappa \frac{\partial \psi}{\partial z} + \Delta_x \psi + \kappa^2 (n^2 - 1)\psi = 0. \tag{3.5}$$

The index of refraction $n(z,x)$ is fluctuating in both the axial z and transversal x variables and thus has the form

$$n^2(z,x) = 1 - 2\sigma V\left(\frac{z}{l_z}, \frac{x}{l_x}\right),$$

where V is a mean-zero random field, and where l_x and l_z are the correlation lengths of V in the transverse and longitudinal directions, respectively. The small parameter σ measures the strength of the fluctuations.

We now introduce two macroscopic distances of wave proapagation: L_x in the x-plane and L_z in the z-direction. We also introduce a carrier wave number κ_0 and replace $\kappa \to \kappa_0\kappa$, κ now being a non-dimensional wavenumber. The physical parameters determined by the medium are the length scales l_x, l_z and the non-dimensional parameter $\sigma \ll 1$.

We present the relationship between the various scalings introduced above that need be satisfied so that wave propagation occurs in a regime close to that of radiative transfer. Equation (3.5) in the non-dimensional variables $z \to z/L_z$, $x \to x/L_x$ becomes

$$\frac{1}{L_z^2}\frac{\partial^2 \psi}{\partial z^2} + \frac{2i\kappa\kappa_0}{L_z}\frac{\partial \psi}{\partial z} + \frac{1}{L_x^2}\Delta_x \psi - 2\kappa^2\kappa_0^2\sigma V\left(\frac{zL_z}{l_z}, \frac{xL_x}{l_x}\right)\psi = 0. \tag{3.6}$$

Let us introduce the following parameters

$$\delta_x = \frac{l_x}{L_x}, \quad \delta_z = \frac{l_z}{L_z}, \quad \gamma_x = \frac{1}{\kappa_0 l_x}, \quad \gamma_z = \frac{1}{\kappa_0 l_z}, \tag{3.7}$$

and recast (3.6) as

$$\gamma_z\delta_z \frac{\partial^2 \psi}{\partial z^2} + 2i\kappa \frac{\partial \psi}{\partial z} + \frac{\delta_x^2\gamma_x^2}{\delta_z\gamma_z}\Delta_x \psi - \frac{2\kappa^2\sigma}{\gamma_z\delta_z} V\left(\frac{z}{\delta_z}, \frac{x}{\delta_x}\right)\psi = 0. \tag{3.8}$$

Let us now assume the following relationships among the various parameters

$$\delta_x = \delta_z \ll 1, \quad \gamma_z = \gamma_x^2 \ll 1, \quad \sigma = \gamma_z\sqrt{\delta_x}, \quad \varepsilon = \delta_x. \tag{3.9}$$

Then (3.8), after multiplication by $\varepsilon/2$, becomes

$$\frac{\gamma_z\varepsilon^2}{2}\frac{\partial^2 \psi}{\partial z^2} + i\kappa\varepsilon \frac{\partial \psi}{\partial z} + \frac{\varepsilon^2}{2}\Delta_x \psi - \kappa^2\sqrt{\varepsilon}V\left(\frac{z}{\varepsilon}, \frac{x}{\varepsilon}\right)\psi = 0. \tag{3.10}$$

We now observe that, when $\kappa = O(1)$ and $\gamma_z \ll 1$, the first term in (3.10) is small and may be neglected in the leading order since $|\varepsilon^2\psi_{zz}| = O(1)$. Then (3.10) becomes

$$i\kappa\varepsilon \frac{\partial \psi}{\partial z} + \frac{\varepsilon^2}{2}\Delta_x \psi - \kappa^2\sqrt{\varepsilon}V\left(\frac{z}{\varepsilon}, \frac{x}{\varepsilon}\right)\psi = 0 \tag{3.11}$$

which is the parabolic wave equation (3.3) in the radiative transfer scaling. The rigorous passage to the parabolic approximation in a three-dimensional layered random medium in a similar scaling is discussed in [1].

Exercise 3.1.1 (i) Show that the above choices imply that

$$l_x \ll l_z.$$

Therefore the correlation length in the longitudinal direction z should be much longer than in the transverse plane x.
(ii) Show that

$$L_x = l_x \frac{l_x^4}{\sigma^2 l_z^4}, \qquad L_z = l_z \frac{l_x^4}{\sigma^2 l_z^4}, \qquad L_x \ll L_z.$$

The latter is the usual constraint for the validity of the parabolic approximation (beam-like structure of the wave).

In the above scalings, there remains one free parameter, namely $\gamma_z = l_z^2/l_x^2$, as one can verify, or equivalently

$$\frac{L_x}{L_z} = \frac{l_x}{l_z} \equiv \varepsilon^{\eta}, \qquad \eta > 0, \tag{3.12}$$

where $\eta > 0$ is necessary since $L_x \ll L_z$. Note that as $\eta \to 0$, we recover an isotropic random medium (with $l_z \equiv l_x$) and the usual regime of radiative transfer. The parabolic (or paraxial) regime thus shares some of the features of the radiative transfer regime, and because the fluctuations depend on the variable z, which plays a similar role to the time variable in the radiative transfer theory, the mathematical analysis is much simplified.

3.2 Wigner Transform and mixture of states

We want to analyze the energy density of the solution to the paraxial wave equation in the limit $\varepsilon \to 0$. Let us recast the above paraxial wave equation as the following Cauchy problem

$$\begin{aligned} & i\varepsilon\kappa \frac{\partial \psi_\varepsilon}{\partial z} + \frac{\varepsilon^2}{2}\Delta\psi_\varepsilon - \kappa^2\sqrt{\varepsilon} V\left(\frac{z}{\varepsilon}, \frac{x}{\varepsilon}\right)\psi_\varepsilon = 0 \\ & \psi_\varepsilon(0,x) = \psi_\varepsilon^0(x;\zeta). \end{aligned} \tag{3.13}$$

Here, the initial data depend on an additional random variable ζ defined over a state space S with a probability measure $d\varpi(\zeta)$. Its use will become clear very soon.

Let us define the Wigner transform as the usual Wigner transform of the field ψ_ε averaged over the space $(S, d\varpi(\zeta))$:

$$W_\varepsilon(z,x,k) = \int_{\mathbb{R}^d \times S} e^{ik\cdot y}\psi_\varepsilon\left(z, x - \frac{\varepsilon y}{2}; \zeta\right)\bar{\psi}_\varepsilon\left(z, x + \frac{\varepsilon y}{2}; \zeta\right)\frac{dy}{(2\pi)^d} d\varpi(\zeta). \tag{3.14}$$

We *assume* that the initial data $W_\varepsilon(0,x,k)$ converges strongly in $L^2(\mathbb{R}^d \times \mathbb{R}^d)$ to a limit $W_0(x,k)$. This is possible thanks to the introduction of a mixture of states, i.e., an integration against the measure $\varpi(d\xi)$. This is the main reason why the space $(S, d\varpi(\zeta))$ is introduced.

Note that the Wigner transform of a *pure* state (e.g. when $\varpi(d\xi)$ concentrates at one point in S) is not even bounded in $L^2(\mathbb{R}^{2d})$ uniformly in ε unless ψ_ε converge strongly to zero in L^2 as $\varepsilon \to 0$. Indeed, if we set

$$W_\varepsilon[u,v](x,k) = \int_{\mathbb{R}^d} e^{ik\cdot y} u_\varepsilon\left(x - \frac{\varepsilon y}{2}\right)\bar{v}_\varepsilon\left(x + \frac{\varepsilon y}{2}\right)\frac{dy}{(2\pi)^d}$$

then we have

$$\int_{\mathbb{R}^{2d}} |W_\varepsilon[u,v]|^2(x,k)\,dx\,dk = \frac{1}{\varepsilon^{2d}}\|u\|^2_{L^2(\mathbb{R}^d)}\|v\|^2_{L^2(\mathbb{R}^d)}. \tag{3.15}$$

We thus need to *regularize* the Wigner transform if we want a uniform bound in $L^2(\mathbb{R}^d\times\mathbb{R}^d)$. This may be achieved by taking random initial data as we did here. We will come back to the effect of not regularizing the Wigner transform at the end of the chapter.

We verify that the Wigner transform satisfies the following evolution equation

$$\frac{\partial W_\varepsilon}{\partial z}+\frac{1}{\kappa}k\cdot\nabla_x W_\varepsilon=\frac{\kappa}{i\sqrt{\varepsilon}}\int_{\mathbb{R}^d}e^{ip\cdot x/\varepsilon}\Big(W_\varepsilon(k-\frac{p}{2})-W_\varepsilon(k+\frac{p}{2})\Big)\frac{d\tilde V\Big(\frac{z}{\varepsilon},p\Big)}{(2\pi)^d}. \tag{3.16}$$

Here, $\tilde V(z,p)$ is the partial Fourier transform of $V(z,x)$ in the variable x. The above evolution equation *preserves* the $L^2(\mathbb{R}^d\times\mathbb{R}^d)$ norm of $W_\varepsilon(t,\cdot,\cdot)$:

Lemma 3.2.1 *Let $W_\varepsilon(t,x,k)$ be the solution of* (3.16) *with initial conditions $W_\varepsilon(0,x,k)$. Then we have*

$$\|W_\varepsilon(t,\cdot,\cdot)\|_{L^2(\mathbb{R}^d\times\mathbb{R}^d)}=\|W_\varepsilon(0,\cdot,\cdot)\|_{L^2(\mathbb{R}^d\times\mathbb{R}^d)},\qquad \textit{for all } t>0. \tag{3.17}$$

Proof. This can be obtained by integrations by parts in (3.16), in a way that is similar to showing that (3.13) preserves the L^2 norm. This can also be obtained from the definition of the Wigner transform in the same way as (3.15) is obtained, and using the preservation of L^2-norm by (3.13). □

3.3 Hypotheses on the randomness

We describe here the construction of the random potential $V(z,x)$. Our main hypothesis is to assume that $V(z,x)$ is a Markov process in the z variable. This is gives us access to a whole machinery relatively similar to the one used in the previous chapter. The Markovian hypothesis is crucial to simplify the mathematical analysis because it allows us to treat the process $z\mapsto(V(z/\varepsilon,x/\varepsilon),W_\varepsilon(z,x,k))$ as jointly Markov.

In addition to being Markovian, $V(z,x)$ is assumed to be stationary in x and z, mean zero, and is constructed in the Fourier space as follows. Let $\mathcal{V}$ be the set of measures of bounded total variation with support inside a ball $B_L=\{|p|\le L\}$

$$\mathcal{V}=\Big\{\hat V:\ \int_{\mathbb{R}^d}|d\hat V|\le C,\ \text{supp}\ \hat V\subset B_L,\ \hat V(p)=\hat V^*(-p)\Big\}, \tag{3.18}$$

and let $\tilde V(z)$ be a mean-zero Markov process on $\mathcal{V}$ with infinitesimal generator Q. The random potential $V(z,x)$ is given by

$$V(z,x)=\int_{\mathbb{R}^d}\frac{d\tilde V(z,p)}{(2\pi)^d}e^{ip\cdot x}. \tag{3.19}$$

It is real-valued and uniformly bounded; $|V(z,x)|\le C$. The correlation function $R(z,x)$ of $V(z,x)$ is

$$R(z,x)=\mathbb{E}\left\{V(s,y)V(z+s,x+y)\right\}\qquad \text{for all } x,y\in\mathbb{R}^d, \text{ and } z,s\in\mathbb{R}. \tag{3.20}$$

In the Fourier domain, this is equivalent to the following statement:

$$\mathbb{E}\Big\{\langle\tilde V(s),\hat\phi\rangle\langle\tilde V(z+s),\hat\psi\rangle\Big\}=(2\pi)^d\int_{\mathbb{R}^d}dp\tilde R(z,p)\hat\phi(p)\hat\psi(-p), \tag{3.21}$$

where $\langle\cdot,\cdot\rangle$ is the usual duality product on $\mathbb{R}^d\times\mathbb{R}^d$, and the power spectrum $\tilde R$ is the Fourier transform of $R(z,x)$ in x:

$$\tilde R(z,p)=\int_{\mathbb{R}^d}dxe^{-ip\cdot x}R(z,x). \tag{3.22}$$

We assume that $\tilde{R}(z,p) \in \mathcal{S}(\mathbb{R}\times\mathbb{R}^d)$, the space of Schwartz functions, for simplicity and define $\hat{R}(\omega,p)$ as

$$\hat{R}(\omega,p) = \int_{\mathbb{R}} dz e^{-i\omega z}\tilde{R}(z,p), \tag{3.23}$$

which is the space-time Fourier transform of R.

We now make additional assumptions on the infinitesimal generator so that the Fredholm alternative holds for the Poisson equation. Namely, we assume that the generator Q is a bounded operator on $L^\infty(\mathcal{V})$ with a unique invariant measure $\pi(\hat{V})$, i.e. a unique normalized measure such that $Q^*\pi = 0$, and assume the existence of a constant $\alpha > 0$ such that if $\langle g, \pi\rangle = 0$, then

$$\|e^{rQ}g\|_{L^\infty_{\mathcal{V}}} \le C\|g\|_{L^\infty_{\mathcal{V}}} e^{-\alpha r}. \tag{3.24}$$

The simplest example of a generator with gap in the spectrum and invariant measure π is a jump process on $\mathcal{V}$ where

$$Qg(\hat{V}) = \int_{\mathcal{V}} g(\hat{V}_1)d\pi(\hat{V}_1) - g(\hat{V}), \qquad \int_{\mathcal{V}} d\pi(\hat{V}) = 1.$$

Given the above hypotheses, the Fredholm alternative holds for the Poisson equation

$$Qf = g, \tag{3.25}$$

provided that g satisfies $\langle \pi, g\rangle = 0$. It has a unique solution f with $\langle \pi, f\rangle = 0$ and $\|f\|_{L^\infty_{\mathcal{V}}} \le C\|g\|_{L^\infty_{\mathcal{V}}}$. The solution f is given explicitly by

$$f(\hat{V}) = -\int_0^\infty dr e^{rQ} g(\hat{V}), \tag{3.26}$$

and the integral converges absolutely thanks to (3.24).

3.4 The Main result

Let us summarize the hypotheses. We define $W_\varepsilon(z,x,k)$ in (3.14) as a mixture of states of solutions to the paraxial wave equation (3.13). We assume that $W_\varepsilon(0,x,k)$ converges *strongly* in $L^2(\mathbb{R}^{2d})$ to its limit $W_0(0,x,k)$. We further assume that the random field $V(z,x)$ satistifies the hypotheses described in Section 3.3. Then we have the following convergence result.

Theorem 3.4.1 *Under the above assumptions, the Wigner distribution W_ε converges in probability and weakly in $L^2(\mathbb{R}^{2d})$ to the solution $\overline{W}$ of the following transport equation*

$$\kappa\frac{\partial\overline{W}}{\partial z} + k\cdot\nabla_x\overline{W} = \kappa^2\mathcal{L}\overline{W}, \tag{3.27}$$

where the scattering kernel has the form

$$\mathcal{L}W(x,k) = \int_{\mathbb{R}^d} \hat{R}\Big(\frac{|p|^2-|k|^2}{2}, p-k\Big)\Big(W(x,p) - W(x,k)\Big)\frac{dp}{(2\pi)^d}. \tag{3.28}$$

More precisely, for any test function $\lambda \in L^2(\mathbb{R}^{2d})$ the process $\langle W_\varepsilon(z),\lambda\rangle$ converges to $\langle\overline{W}(z),\lambda\rangle$ in probability as $\varepsilon\to 0$, uniformly on finite intervals $0\le z\le Z$.

Note that the whole process W_ε, and not only its average $\mathbb{E}\{W_\varepsilon\}$ converges to the (deterministic) limit $\overline{W}$. This means that the process W_ε is *statistically stable* in the limit $\varepsilon \to 0$. The process $W_\varepsilon(z,x,k)$ does not converge pointwise to the deterministic limit: averaging against a test function $\lambda(x,k)$ is necessary.

The next section is devoted to a proof of the theorem. We first summarize the main ingredients of the proof. Recall that the main mathematical assumption is that $V(z,x)$ is Markov in the z variable. Let us set $L > 0$ and consider $z \in [0,L]$. Then $(V(z/\varepsilon, x/\varepsilon), W_\varepsilon(z,x,k))$ is jointly Markov in the space $\mathcal{V} \times \mathcal{X}$, where $\mathcal{X} = \mathcal{C}([0,L]; B_W)$, where $B_W = \{\|W\|_2 \le C\}$ is an appropriate ball in $L^2(\mathbb{R}^d \times \mathbb{R}^d)$.

Step 1. Evolution equation and random process. Since κ plays no role in the derivation, we set $\kappa = 1$ for simplicity. Recall that W_ε satisfies the Cauchy problem

$$\frac{\partial W_\varepsilon}{\partial z} + k \cdot \nabla_x W_\varepsilon = \mathcal{L}_\varepsilon W_\varepsilon,$$

with $W_\varepsilon(0,x,k) = W_\varepsilon^0(x,k)$, where

$$\mathcal{L}_\varepsilon W_\varepsilon = \frac{1}{i\sqrt{\varepsilon}} \int_{\mathbb{R}^d} \frac{d\tilde{V}(\frac{z}{\varepsilon}, p)}{(2\pi)^d} e^{ip\cdot x/\varepsilon} \left[W_\varepsilon(x, k - \frac{p}{2}) - W_\varepsilon(x, k + \frac{p}{2}) \right]. \tag{3.29}$$

The solution to the above Cauchy problem is understood in the sense that for every smooth test function $\lambda(z,x,k)$, we have

$$\langle W_\varepsilon(z), \lambda(z)\rangle - \langle W_\varepsilon^0, \lambda(0)\rangle = \int_0^z \langle W_\varepsilon(s), \left(\frac{\partial}{\partial s} + k\cdot\nabla_x + \mathcal{L}_\varepsilon\right)\lambda(s)\rangle ds.$$

Here, we have used that $\mathcal{L}_\varepsilon$ is an anti-self-adjoint operator for $\langle\cdot,\cdot\rangle$. Therefore, for a smooth function $\lambda_0(x,k)$, we obtain $\langle W_\varepsilon(z), \lambda_0\rangle = \langle W_\varepsilon^0, \lambda_\varepsilon(0)\rangle$, where $\lambda_\varepsilon(s)$ is the solution of the backward problem

$$\frac{\partial \lambda_\varepsilon}{\partial s} + k\cdot\nabla_x \lambda_\varepsilon + \mathcal{L}_\varepsilon \lambda_\varepsilon(s) = 0, \quad 0 \le s \le z,$$

with the terminal condition $\lambda_\varepsilon(z,x,k) = \lambda_0(x,k)$.

Step 2. Tightness of the family of $\varepsilon-$measures. The above construction defines the process $W_\varepsilon(z)$ in $L^2(\mathbb{R}^{2d})$ and generates a corresponding measure P_ε on the state space $C([0,L]; L^2(\mathbb{R}^{2d}))$ of continuous functions in time with values in L^2. The measure P_ε is actually supported on paths inside $\mathcal{X}$ defined above, which is the state space for the random process $W_\varepsilon(z)$. With its natural topology and the Borel $\sigma-$algebra $\mathcal{F}$, $(\mathcal{X}, \mathcal{F}, P_\varepsilon)$ defines a probability space on which $W_\varepsilon(z)$ is a random variable. Then $\mathcal{F}_s$ is defined as the filtration of the process $W_\varepsilon(z)$, that is, the filtration generated by $\{W_\varepsilon(\tau), \tau < s\}$.

The family P_ε parameterized by $\varepsilon_0 > \varepsilon > 0$ will be shown to be *tight*. This in turns implies that P_ε converges weakly to a limit probability measure P. This means that we can extract a subsequence of P_ε, still denoted by P_ε, such that for all bounded continuous functions f defined on $\mathcal{X}$, we have

$$\mathbb{E}^{P_\varepsilon}\{f\} \equiv \int_{\mathcal{X}} f(\omega) dP_\varepsilon(\omega) \to \int_{\mathcal{X}} f(\omega) dP(\omega) \equiv \mathbb{E}^P\{f\}, \text{ as } \quad \varepsilon \to 0. \tag{3.30}$$

Step 3. Construction of a first approximate martingale. Once tightness is ensured, the proof of convergence of W_ε to its deterministic limit is obtained in two steps. Let us fix a deterministic test function $\lambda(z, x, k)$. We use the Markovian property of the random field $V(z, x)$ in z to construct a first functional $G_\lambda : \mathcal{X} \to C[0, L]$ by

$$G_\lambda[W](z) = \langle W, \lambda\rangle(z) - \int_0^z \langle W, \frac{\partial\lambda}{\partial z} + k \cdot \nabla_x\lambda + \mathcal{L}\lambda\rangle(\zeta)d\zeta. \tag{3.31}$$

Here, $\mathcal{L}$ is the limiting scattering kernel defined in (3.28). We will show that G_λ is an approximate P_ε-martingale (with respect to the filtration $\mathcal{F}_s$), and more precisely that

$$\left|\mathbb{E}^{P_\varepsilon}\left\{G_\lambda[W](z)|\mathcal{F}_s\right\} - G_\lambda[W](s)\right| \le C_{\lambda,L}\sqrt{\varepsilon} \tag{3.32}$$

uniformly for all $W \in \mathcal{X}$ and $0 \le s < z \le L$. The two convergences (3.30) and (3.32) (weak against strong) show that

$$\mathbb{E}^P\left\{G_\lambda[W](z)|\mathcal{F}_s\right\} - G_\lambda[W](s) = 0, \tag{3.33}$$

which implies that $G_\lambda[W]$ so constructed is a P-martingale. Taking $s = 0$ above, we obtain the transport equation (3.27) for $\overline{W} = \mathbb{E}^P\{W(z)\}$ in its weak formulation.

Step 4. Construction of a second approximate martingale and convergence of the full family of $\varepsilon-$measures. So far, we have characterized the convergence of the first moment of P_ε. We now consider the convergence of the second moment and show that the variance of the limiting process vanishes, whence the convergence to a deterministic process.

We will show that for every test function $\lambda(z, x, k)$, the new functional

$$G_{2,\lambda}[W](z) = \langle W, \lambda\rangle^2(z) - 2\int_0^z \langle W, \lambda\rangle(\zeta)\langle W, \frac{\partial\lambda}{\partial z} + k \cdot \nabla_x\lambda + \mathcal{L}\lambda\rangle(\zeta)d\zeta \tag{3.34}$$

is also an approximate P_ε-martingale. We then obtain that

$$\mathbb{E}^{P_\varepsilon}\left\{\langle W, \lambda\rangle^2\right\} \to \langle\overline{W}, \lambda\rangle^2. \tag{3.35}$$

This crucial convergence implies convergence in probability. It follows that the limit measure P is unique and deterministic, and that the whole sequence P_ε converges.

3.5 Proof of Theorem 3.4.1

The proof of tightness of the family of measures P_ε is postponed to the end of the section as it requires estimates that are developed in the proofs of convergence of the approximate martingales. We thus start with the latter proofs.

3.5.1 Convergence in expectation

To obtain the approximate martingale property (3.32), one has to consider the conditional expectation of functionals $F(W, \hat{V})$ with respect to the probability measure $\tilde{P}_\varepsilon$ on the space $C([0, L]; \mathcal{V} \times B_W)$ generated by $V(z/\varepsilon)$ and the Cauchy problem (3.16). The only functions we need consider are in fact of the form $F(W, \hat{V}) = \langle W, \lambda(\hat{V})\rangle$ with $\lambda \in L^\infty(\mathcal{V}; C^1([0, L]; \mathcal{S}(\mathbb{R}^{2d})))$. Given a function $F(W, \hat{V})$ let us define the conditional expectation

$$\mathbb{E}^{\tilde{P}_\varepsilon}_{W,\hat{V},z}\left\{F(W, \hat{V})\right\}(\tau) = \mathbb{E}^{\tilde{P}_\varepsilon}\left\{F(W(\tau), \tilde{V}(\tau))|\ W(z) = W, \tilde{V}(z) = \hat{V}\right\}, \quad \tau \ge z.$$

The weak form of the infinitesimal generator of the Markov process generated by $\tilde{P}_\varepsilon$ is then given by

$$\frac{d}{dh}\mathbb{E}^{\tilde{P}_\varepsilon}_{W,\hat{V},z}\left\{\langle W,\lambda(\hat{V})\rangle\right\}(z+h)\Big|_{h=0} = \frac{1}{\varepsilon}\langle W,Q\lambda\rangle + \left\langle W,\left(\frac{\partial}{\partial z}+k\cdot\nabla_x+\frac{1}{\sqrt{\varepsilon}}\mathcal{K}[\hat{V},\frac{x}{\varepsilon}]\right)\lambda\right\rangle. \tag{3.36}$$

where the operator $\mathcal{K}$ is defined by

$$\mathcal{K}[\hat{V},\eta]\psi(x,\eta,k,\hat{V}) = \frac{1}{i}\int_{\mathbb{R}^d}\frac{d\hat{V}(p)}{(2\pi)^d}e^{ip\cdot\eta}\left[\psi(x,\eta,k-\frac{p}{2})-\psi(x,\eta,k+\frac{p}{2})\right]. \tag{3.37}$$

The above equality implies that

$$G^\varepsilon_\lambda = \langle W,\lambda(\hat{V})\rangle(z) - \int_0^z\left\langle W,\left(\frac{1}{\varepsilon}Q+\frac{\partial}{\partial z}+k\cdot\nabla_x+\frac{1}{\sqrt{\varepsilon}}\mathcal{K}[\hat{V},\frac{x}{\varepsilon}]\right)\lambda\right\rangle(s)ds \tag{3.38}$$

is a $\tilde{P}_\varepsilon$-martingale since the drift term has been subtracted.

Given a test function $\lambda(z,x,k)\in C^1([0,L];\mathcal{S})$ we construct a function

$$\lambda_\varepsilon(z,x,k,\hat{V}) = \lambda(z,x,k)+\sqrt{\varepsilon}\lambda_1^\varepsilon(z,x,k,\hat{V})+\varepsilon\lambda_2^\varepsilon(z,x,k,\hat{V}), \tag{3.39}$$

with $\lambda^\varepsilon_{1,2}(t)$ bounded in $L^\infty(\mathcal{V};L^2(\mathbb{R}^{2d}))$ uniformly in $z\in[0,L]$. This is the method of **perturbed test function**. Rather than performing asymptotic expansions on the Wigner transform itself, which is not sufficiently smooth to justify Taylor expansions, we perform the expansion on smooth test functions.

The functions $\lambda^\varepsilon_{1,2}$ will be chosen to remove all high-order terms in the definition of the martingale (3.38), i.e., so that

$$\|G^\varepsilon_{\lambda_\varepsilon}(z)-G_\lambda(z)\|_{L^\infty(\mathcal{V})}\le C_\lambda\sqrt{\varepsilon} \tag{3.40}$$

for all $z\in[0,L]$. Here $G^\varepsilon_{\lambda_\varepsilon}$ is defined by (3.38) with λ replaced by λ_ε, and G_λ is defined by (3.31). The approximate martingale property (3.32) follows from this.

The functions λ_1^ε and λ_2^ε are as follows. Let $\lambda_1(z,x,\eta,k,\hat{V})$ be the mean-zero solution of the Poisson equation

$$k\cdot\nabla_\eta\lambda_1+Q\lambda_1=-\mathcal{K}\lambda. \tag{3.41}$$

It is given explicitly by

$$\lambda_1(z,x,\eta,k,\hat{V}) = \frac{1}{i}\int_0^\infty dre^{rQ}\int_{\mathbb{R}^d}\frac{d\hat{V}(p)}{(2\pi)^d}e^{ir(k\cdot p)+i(\eta\cdot p)}\left[\lambda(z,x,k-\frac{p}{2})-\lambda(z,x,k+\frac{p}{2})\right]. \tag{3.42}$$

Then we let $\lambda_1^\varepsilon(z,x,k,\hat{V})=\lambda_1(z,x,x/\varepsilon,k,\hat{V})$. Furthermore, the second order corrector is given by $\lambda_2^\varepsilon(z,x,k,\hat{V})=\lambda_2(z,x,x/\varepsilon,k,\hat{V})$ where $\lambda_2(z,x,\eta,k,\hat{V})$ is the mean-zero solution of

$$k\cdot\nabla_\eta\lambda_2+Q\lambda_2=\mathcal{L}\lambda-\mathcal{K}\lambda_1, \tag{3.43}$$

which exists because $\mathbb{E}\left\{\mathcal{K}\lambda_1\right\}=\mathcal{L}\lambda$, and is given by

$$\lambda_2(z,x,\eta,k,\hat{V}) = -\int_0^\infty dre^{rQ}\left[\mathcal{L}\lambda(z,x,k)-[\mathcal{K}\lambda_1](z,x,\eta+rk,k,\hat{V})\right].$$

Using (3.41) and (3.43) we have

$$\begin{aligned}\frac{d}{dh}\mathbb{E}^{\tilde{P}_\varepsilon}_{W,\hat{V},z}\left\{\langle W,\lambda_\varepsilon\rangle\right\}(z+h)\Big|_{h=0} &= \left\langle W,\left(\frac{\partial}{\partial z}+k\cdot\nabla_x+\frac{1}{\sqrt{\varepsilon}}\mathcal{K}[\hat{V},\frac{x}{\varepsilon}]+\frac{1}{\varepsilon}Q\right)\left(\lambda+\sqrt{\varepsilon}\lambda_1^\varepsilon+\varepsilon\lambda_2^\varepsilon\right)\right\rangle\\ &= \left\langle W,\left(\frac{\partial}{\partial z}+k\cdot\nabla_x\right)\lambda+\mathcal{L}\lambda\right\rangle+\left\langle W,\left(\frac{\partial}{\partial z}+k\cdot\nabla_x\right)\left(\sqrt{\varepsilon}\lambda_1^\varepsilon+\varepsilon\lambda_2^\varepsilon\right)+\sqrt{\varepsilon}\mathcal{K}[\hat{V},\frac{x}{\varepsilon}]\lambda_2^\varepsilon\right\rangle\\ &= \left\langle W,\left(\frac{\partial}{\partial z}+k\cdot\nabla_x\right)\lambda+\mathcal{L}\lambda\right\rangle+\sqrt{\varepsilon}\langle W,\zeta_\varepsilon^\lambda\rangle,\end{aligned}$$

with

$$\zeta_\varepsilon^\lambda = \left(\frac{\partial}{\partial z} + k\cdot\nabla_x\right)\lambda_1^\varepsilon + \sqrt{\varepsilon}\left(\frac{\partial}{\partial z} + k\cdot\nabla_x\right)\lambda_2^\varepsilon + \mathcal{K}[\hat{V},\frac{x}{\varepsilon}]\lambda_2^\varepsilon.$$

The terms $k\cdot\nabla_x\lambda_{1,2}^\varepsilon$ above are understood as differentiation with respect to the slow variable x only, and not with respect to $\eta = x/\varepsilon$. It follows that $G_{\lambda_\varepsilon}^\varepsilon$ is given by

$$G_{\lambda_\varepsilon}^\varepsilon(z) = \langle W(z),\lambda_\varepsilon\rangle - \int_0^z ds\left\langle W,\left(\frac{\partial}{\partial z} + k\cdot\nabla_x + \mathcal{L}\right)\lambda\right\rangle(s) - \sqrt{\varepsilon}\int_0^z ds\langle W,\zeta_\varepsilon^\lambda\rangle(s) \quad (3.44)$$

and is a martingale with respect to the measure $\tilde{P}_\varepsilon$ defined on $D([0,L];X\times\mathcal{V})$, the space of right-continuous paths with left-side limits [11]. The estimate (3.32) follows from the following two lemmas.

Lemma 3.5.1 *Let $\lambda \in C^1([0,L];\mathcal{S}(\mathbb{R}^{2d}))$. Then there exists a constant $C_\lambda > 0$ independent of $z\in[0,L]$ so that the correctors $\lambda_1^\varepsilon(z)$ and $\lambda_2^\varepsilon(z)$ satisfy the uniform bounds*

$$\|\lambda_1^\varepsilon(z)\|_{L^\infty(\mathcal{V};L^2)} + \|\lambda_2^\varepsilon(z)\|_{L^\infty(\mathcal{V};L^2)} \le C_\lambda \quad (3.45)$$

and

$$\left\|\frac{\partial\lambda_1^\varepsilon(z)}{\partial z} + k\cdot\nabla_x\lambda_1^\varepsilon(z)\right\|_{L^\infty(\mathcal{V};L^2)} + \left\|\frac{\partial\lambda_2^\varepsilon(z)}{\partial z} + k\cdot\nabla_x\lambda_2^\varepsilon(z)\right\|_{L^\infty(\mathcal{V};L^2)} \le C_\lambda. \quad (3.46)$$

Lemma 3.5.2 *There exists a constant C_λ such that*

$$\|\mathcal{K}[\hat{V},x/\varepsilon]\|_{L^2\to L^2} \le C$$

for any $\hat{V}\in\mathcal{V}$ and all $\varepsilon\in(0,1]$.

Indeed, (3.45) implies that $|\langle W,\lambda\rangle - \langle W,\lambda_\varepsilon\rangle| \le C\sqrt{\varepsilon}$ for all $W\in X$ and $\hat{V}\in\mathcal{V}$, while (3.46) and Lemma 3.5.2 imply that for all $z\in[0,L]$

$$\|\zeta_\varepsilon^\lambda(z)\|_{L^2} \le C, \quad (3.47)$$

for all $\hat{V}\in\mathcal{V}$ so that (3.32) follows.

Proof of Lemma 3.5.2. Lemma 3.5.2 follows immediately from the definition of $\mathcal{K}$, the bound (3.18) and the Cauchy-Schwarz inequality.

We now prove Lemma 3.5.1. We will omit the z-dependence of the test function λ to simplify the notation.

Proof of Lemma 3.5.1. We only prove (3.45). Since $\lambda\in\mathcal{S}(\mathbb{R}^{2d})$, there exists a constant C_λ so that

$$|\lambda(x,k)| \le \frac{C_\lambda}{(1+|x|^{5d})(1+|k|^{5d})}.$$

The value of the exponents $5d$ is by no means optimal, and is sufficient in what follows. Then we obtain using (3.18) and (3.24)

$$\begin{aligned}
|\lambda_1^\varepsilon(z,x,k,\hat{V})| &= C\left|\int_0^\infty dre^{rQ}\int_{\mathbb{R}^d} d\hat{V}(p)e^{ir(k\cdot p)+i(x\cdot p)/\varepsilon}\left[\lambda(z,x,k-\frac{p}{2}) - \lambda(z,x,k+\frac{p}{2})\right]\right| \\
&\le C\int_0^\infty dre^{-\alpha r}\sup_{\hat{V}}\int_{\mathbb{R}^d}|d\hat{V}(p)|\left[|\lambda(z,x,k-\frac{p}{2})| + |\lambda(z,x,k+\frac{p}{2})|\right] \\
&\le \frac{C}{(1+|x|^{5d})(1+(|k|-L)^{5d}\chi_{|k|\ge 5L}(k))},
\end{aligned}$$

and the L^2-bound on λ_1 follows.

We show next that λ_2^ε is uniformly bounded. We have

$$\lambda_2^\varepsilon(x,k,\hat V) = -\int_0^\infty dre^{rQ}\Big[\mathcal{L}\lambda(x,k) - \frac{1}{i}\int_{\mathbb{R}^d}\frac{d\hat V(p)}{(2\pi)^d}e^{ip\cdot(x/\varepsilon+rk)}$$
$$\times\Big[\lambda_1(x,\frac{x}{\varepsilon}+rk,k-\frac{p}{2},\hat V) - \lambda_1(x,\frac{x}{\varepsilon}+rk,k+\frac{p}{2},\hat V)\Big]\Big].$$

The second term above may be written as

$$\frac{1}{i}\int_{\mathbb{R}^d}\frac{d\hat V(p)}{(2\pi)^d}e^{ip\cdot(x/\varepsilon+rk)}\Big[\lambda_1(x,\frac{x}{\varepsilon}+rk,k-\frac{p}{2},\hat V) - \lambda_1(x,\frac{x}{\varepsilon}+rk,k+\frac{p}{2},\hat V)\Big]$$
$$= -\int_{\mathbb{R}^d}\frac{d\hat V(p)}{(2\pi)^d}e^{ip\cdot(x/\varepsilon+rk)}\int_0^\infty dse^{sQ}\int_{\mathbb{R}^d}\frac{d\hat V(q)}{(2\pi)^d}e^{is(k-p/2)\cdot q+i(x/\varepsilon+rk)\cdot q}$$
$$\times\Big[\lambda(x,k-\frac{p}{2}-\frac{q}{2}) - \lambda(x,k-\frac{p}{2}+\frac{q}{2})\Big]$$
$$+\int_{\mathbb{R}^d}\frac{d\hat V(p)}{(2\pi)^d}e^{ip\cdot(x/\varepsilon+rk)}\int_0^\infty dse^{sQ}\int_{\mathbb{R}^d}\frac{d\hat V(q)}{(2\pi)^d}e^{is(k+p/2)\cdot q+i(x/\varepsilon+rk)\cdot q}$$
$$\times\Big[\lambda(x,k+\frac{p}{2}-\frac{q}{2}) - \lambda(x,k+\frac{p}{2}+\frac{q}{2})\Big].$$

Therefore we obtain

$$|\lambda_2^\varepsilon(x,k,\hat V)| \le C\int_0^\infty dre^{-\alpha r}\Big[|\mathcal{L}\lambda(x,k)| + \sup_{\hat V}\int_{\mathbb{R}^d}|d\hat V(p)|\int_0^\infty dse^{-\alpha s}\sup_{\hat V_1}\int_{\mathbb{R}^d}|d\hat V_1(q)|$$
$$\times\Big(|\lambda(x,k-\frac{p}{2}-\frac{q}{2})| + |\lambda(x,k-\frac{p}{2}+\frac{q}{2})| + |\lambda(x,k+\frac{p}{2}-\frac{q}{2})| + \lambda(x,k+\frac{p}{2}+\frac{q}{2})\Big)\Big]$$
$$\le C\left[|\mathcal{L}\lambda(x,k)| + \frac{1}{(1+|x|^{5d})(1+(|k|-L)^{5d}\chi_{|k|\ge 5L}(k))}\right],$$

and the L^2-bound on λ_2^ε in (3.45) follows because the operator $\mathcal{L}: L^2 \to L^2$ is bounded. The proof of (3.46) is very similar and is left as a painful exercise.

Lemma 3.5.1 and Lemma 3.5.2 together with (3.44) imply the bound (3.40). The tightness of measures P_ε given by Lemma 3.5.4 implies then that the expectation $\mathbb{E}\{W_\varepsilon(z,x,k)\}$ converges weakly in $L^2(\mathbb{R}^{2d})$ to the solution $\overline{W}(z,x,k)$ of the transport equation for each $z \in [0,L]$.

3.5.2 Convergence in probability

We now prove that for any test function λ the second moment $\mathbb{E}\{\langle W_\varepsilon,\lambda\rangle^2\}$ converges to $\langle\overline{W},\lambda\rangle^2$. This will imply the convergence in probability claimed in Theorem 3.4.1. The proof is similar to that for $\mathbb{E}\{\langle W_\varepsilon,\lambda\rangle\}$ and is based on constructing an appropriate approximate martingale for the functional $\langle W\otimes W,\mu\rangle$, where $\mu(z,x_1,k_1,x_2,k_2)$ is a test function, and $W\otimes W(z,x_1,k_1,x_2,k_2) = W(z,x_1,k_1)W(z,x_2,k_2)$. We need to consider the action of the infinitesimal generator on functions of W and $\hat V$ of the form

$$F(W,\hat V) = \langle W(x_1,k_1)W(x_2,k_2),\mu(z,x_1,k_1,x_2,k_2,\hat V)\rangle = \langle W\otimes W,\mu(\hat V)\rangle$$

where μ is a given function. The infinitesimal generator acts on such functions as

$$\frac{d}{dh}\mathbb{E}^{\tilde P_\varepsilon}_{W,\hat V,z}\Big\{\langle W\otimes W,\mu(\hat V)\rangle\Big\}(z+h)\Big|_{h=0} = \frac{1}{\varepsilon}\langle W\otimes W,Q\lambda\rangle + \langle W\otimes W,\mathcal{H}_2^\varepsilon\mu\rangle, \quad (3.48)$$

where

$$\mathcal{H}_2^\varepsilon \mu = \sum_{j=1}^{2} \frac{1}{\sqrt{\varepsilon}} \mathcal{K}_j \left[\hat{V}, \frac{x^j}{\varepsilon}\right] \mu + k^j \cdot \nabla_{x^j} \mu, \tag{3.49}$$

with

$$\mathcal{K}_1[\hat{V}, \eta_1]\mu = \frac{1}{i} \int_{\mathbb{R}^d} d\hat{V}(p) e^{i(p\cdot\eta_1)} \left[\mu(k_1 - \frac{p}{2}, k_2) - \mu(k_1 + \frac{p}{2}, k_2)\right]$$

and

$$\mathcal{K}_2[\hat{V}, \eta_2]\mu = \frac{1}{i} \int_{\mathbb{R}^d} d\hat{V}(p) e^{i(p\cdot\eta_2)} \left[\mu(k_1, k_2 - \frac{p}{2}) - \mu(k_1, k_2 + \frac{p}{2})\right].$$

Therefore the functional

$$\begin{aligned} &G_\mu^{2,\varepsilon} = \langle W \otimes W, \mu(\hat{V})\rangle(z) \\ &- \int_0^z \left\langle W \otimes W, \Big(\frac{1}{\varepsilon} Q + \frac{\partial}{\partial z} + k_1 \cdot \nabla_{x_1} + k_2 \cdot \nabla_{x_2} + \frac{1}{\sqrt{\varepsilon}}(\mathcal{K}_1[\hat{V}, \frac{x_1}{\varepsilon}] + \mathcal{K}_2[\hat{V}, \frac{x_2}{\varepsilon}])\Big)\mu \right\rangle (s) ds \end{aligned} \tag{3.50}$$

is a $\tilde{P}^\varepsilon$ martingale. We let $\mu(z, x, \mathbf{K}) \in \mathcal{S}(\mathbb{R}^{2d} \times \mathbb{R}^{2d})$ be a test function independent of $\hat{V}$, where $x = (x_1, x_2)$, and $\mathbf{K} = (k_1, k_2)$. We define an approximation

$$\mu_\varepsilon(z, x, \mathbf{K}) = \mu(z, x, \mathbf{K}) + \sqrt{\varepsilon}\mu_1(z, x, x/\varepsilon, \mathbf{K}) + \varepsilon\mu_2(x, x/\varepsilon, \mathbf{K}).$$

We will use the notation $\mu_1^\varepsilon(z, x, \mathbf{K}) = \mu_1(z, x, x/\varepsilon, \mathbf{K})$ and $\mu_2^\varepsilon(z, x, \mathbf{K}) = \mu_2(z, x, x/\varepsilon, \mathbf{K})$. The functions μ_1 and μ_2 are to be determined. We now use (3.48) to get

$$\begin{aligned} D_\varepsilon := \frac{d}{dh}\Big|_{h=0} &\mathbb{E}_{W,\hat{V},z}(\langle W \otimes W, \mu_\varepsilon(\hat{V})\rangle)(z+h) = \frac{1}{\varepsilon} \left\langle W \otimes W, \left(Q + \sum_{j=1}^{2} k^j \cdot \nabla_{\eta^j}\right) \mu \right\rangle \\ &+ \frac{1}{\sqrt{\varepsilon}} \left\langle W \otimes W, \left(Q + \sum_{j=1}^{2} k^j \cdot \nabla_{\eta^j}\right) \mu_1 + \sum_{j=1}^{2} \mathcal{K}_j \left[\hat{V}, \eta^j\right] \mu \right\rangle \\ &+ \left\langle W \otimes W, \left(Q + \sum_{j=1}^{2} k^j \cdot \nabla_{\eta^j}\right) \mu_2 + \sum_{j=1}^{2} \mathcal{K}_j \left[\hat{V}, \eta^j\right] \mu_1 + \frac{\partial \mu}{\partial z} + \sum_{j=1}^{2} k^j \cdot \nabla_{x^j} \mu \right\rangle \\ &+ \sqrt{\varepsilon} \left\langle W \otimes W, \sum_{j=1}^{2} \mathcal{K}_j \left[\hat{V}, \eta^j\right] \mu_2 + \left(\frac{\partial}{\partial z} + \sum_{j=1}^{2} k^j \cdot \nabla_{x^j}\right) (\mu_1 + \sqrt{\varepsilon}\mu_2) \right\rangle. \end{aligned} \tag{3.51}$$

The above expression is evaluated at $\eta_j = x_j/\varepsilon$. The term of order ε^{-1} in D_ε vanishes since μ is independent of V and the fast variable η. We cancel the term of order $\varepsilon^{-1/2}$ in the same way as before by defining μ_1 as the unique mean-zero (in the variables $\hat{V}$ and $\eta = (\eta_1, \eta_2)$) solution of

$$(Q + \sum_{j=1}^{2} k^j \cdot \nabla_{\eta^j})\mu_1 + \sum_{j=1}^{2} \mathcal{K}_j \left[\hat{V}, \eta^j\right] \mu = 0. \tag{3.52}$$

It is given explicitly by

$$\begin{aligned} \mu_1(x, \eta, \mathbf{K}, \hat{V}) = &\frac{1}{i} \int_0^\infty dr e^{rQ} \int_{\mathbb{R}^d} d\hat{V}(p) e^{ir(k_1\cdot p) + i(\eta_1 \cdot p)} \left[\mu(k_1 - \frac{p}{2}, k_2) - \mu(k_1 + \frac{p}{2}, k_2)\right] \\ &+ \frac{1}{i} \int_0^\infty dr e^{rQ} \int_{\mathbb{R}^d} d\hat{V}(p) e^{ir(k_2\cdot p) + i(\eta_2 \cdot p)} \left[\mu(k_1, k_2 - \frac{p}{2}) - \mu(k_1, k_2 + \frac{p}{2})\right]. \end{aligned}$$

When μ has the form $\mu = \lambda\otimes\lambda$, then μ_1 has the form $\mu_1 = \lambda_1\otimes\lambda+\lambda\otimes\lambda_1$ with the corrector λ_1 given by (3.42). Let us also define μ_2 as the mean zero with respect to π_V solution of

$$(Q+\sum_{j=1}^{2} k^j\cdot\nabla_{\eta^j})\mu_2+\sum_{j=1}^{2}\mathcal{K}_j\Big[\hat V,\eta^j\Big]\mu_1=\overline{\sum_{j=1}^{2}\mathcal{K}_j\Big[\hat V,\eta^j\Big]\mu_1}, \tag{3.53}$$

where $\overline{f}=\int d\pi_V f$. The function μ_2 is given by

$$\begin{aligned}\mu_2(x,\eta,\mathbf{K},\hat V) = &-\int_0^\infty dr e^{rQ}\big[\overline{\mathcal{K}_1[\hat V,\eta_1+rk_1]\mu_1(x,\eta+r\mathbf{K},\mathbf{K})}\\ &-[\mathcal{K}_1[\hat V,\eta_1+rk_1]\mu_1](x,\eta+r\mathbf{K},\mathbf{K},\hat V)\big]\\ &-\int_0^\infty dr e^{rQ}\big[\overline{\mathcal{K}_2[\hat V,k_2+r\eta_2]\mu_1(x,\eta+r\mathbf{K},\mathbf{K})}\\ &-[\mathcal{K}_2[\hat V,\eta_2+rk_2]\mu_1](x,\eta+r\mathbf{K},\mathbf{K},\hat V)\big].\end{aligned} \tag{3.54}$$

Unlike the first corrector μ_1, the second corrector μ_2 may not be written as an explicit sum of tensor products even if μ has the form $\mu=\lambda\otimes\lambda$ because μ_1 depends on $\hat V$.

The $\tilde P^\varepsilon$-martingale $G^{2,\varepsilon}_{\mu_\varepsilon}$ is given by

$$\begin{aligned}G^{2,\varepsilon}_{\mu} = &\langle W\otimes W,\mu(\hat V)\rangle(z)-\int_0^z\Big\langle W\otimes W,\Big(\frac{\partial}{\partial z}+k_1\cdot\nabla_{x_1}+k_2\cdot\nabla_{x_2}+\mathcal{L}_2^\varepsilon\Big)\mu\Big\rangle(s)ds\\ &-\sqrt{\varepsilon}\int_0^z\langle W\otimes W,\zeta_\varepsilon^\mu\rangle(s)ds,\end{aligned} \tag{3.55}$$

where ζ_ε^μ is given by

$$\zeta^\varepsilon_\mu=\sum_{j=1}^2\mathcal{K}_j\left[\hat V,\frac{x_j}{\varepsilon}\right]\mu_2^\varepsilon+\left(\frac{\partial}{\partial z}+\sum_{j=1}^2 k^j\cdot\nabla_{x^j}\right)(\mu_1^\varepsilon+\sqrt{\varepsilon}\mu_2^\varepsilon)$$

and the operator $\mathcal{L}_2^\varepsilon$ is defined by

$$\begin{aligned}\mathcal{L}_2^\varepsilon\mu = &-\frac{1}{(2\pi)^d}\int_0^\infty dr\int_{\mathbb{R}^d}dp\tilde R(r,p)\Big[e^{ir(k_1+\frac{p}{2})\cdot p}(\mu(z,x_1,k_1,x_2,k_2)-\mu(z,x_1,k_1+p,x_2,k_2))\\ &-e^{ir(k_1-\frac{p}{2})\cdot p}(\mu(z,x_1,k_1-p,x_2,k_2)-\mu(z,x_1,k_1,x_2,k_2))\Big]\\ &+\Big[e^{ip\cdot\frac{x_2-x_1}{\varepsilon}+irk_2\cdot p}(\mu(z,x_1,k_1+\tfrac{p}{2},x_2,k_2-\tfrac{p}{2})-\mu(z,x_1,k_1+\tfrac{p}{2},x_2,k_2+\tfrac{p}{2}))\\ &-e^{ip\cdot\frac{x_2-x_1}{\varepsilon}+irk_2\cdot p}(\mu(z,x_1,k_1-\tfrac{p}{2},x_2,k_2-\tfrac{p}{2})-\mu(z,x_1,k_1-\tfrac{p}{2},x_2,k_2+\tfrac{p}{2}))\Big]\\ &+\Big[e^{irk_1\cdot p+i\frac{x_1-x_2}{\varepsilon}\cdot p}(\mu(z,x_1,k_1-\tfrac{p}{2},x_2,k_2+\tfrac{p}{2})-\mu(z,x_1,k_1-\tfrac{p}{2},x_2,k_2-\tfrac{p}{2}))\\ &-e^{irk_1\cdot p+i\frac{x_1-x_2}{\varepsilon}\cdot p}(\mu(z,x_1,k_1+\tfrac{p}{2},x_2,k_2+\tfrac{p}{2})-\mu(z,x_1,k_1+\tfrac{p}{2},x_2,k_2-\tfrac{p}{2}))\Big]\\ &+\Big[e^{ir(k_2+\frac{p}{2})\cdot p}(\mu(z,x_1,k_1,x_2,k_2)-\mu(z,x_1,k_1,x_2,k_2+p))\\ &-e^{ir(k_2-\frac{p}{2})\cdot p}(\mu(z,x_1,k_1,x_2,k_2-p)-\mu(z,x_1,k_1,x_2,k_2))\Big].\end{aligned} \tag{3.56}$$

We have used in the calculation of $\mathcal{L}_2^\varepsilon$ that for a sufficiently regular function f, we have

$$\mathbb{E}\left[\int_{\mathbb{R}^d}\frac{d\hat V(q)}{(2\pi)^d}\int_0^\infty dr\,e^{rQ}\int_{\mathbb{R}^d}d\hat V(p)f(r,p,q)\right]=\int_0^\infty dr\int_{\mathbb{R}^d}\tilde R(r,p)f(r,p,-p)dp.$$

The bound on ζ_ε^μ is similar to that on $\zeta_\varepsilon^\lambda$ obtained previously as the correctors μ_j^ε satisfy the same kind of estimates as the correctors λ_j:

Lemma 3.5.3 *There exists a constant $C_\mu > 0$ so that the functions $\mu_{1,2}^\varepsilon$ obey the uniform bounds*

$$\|\mu_1^\varepsilon(z)\|_{L^2(\mathbb{R}^{2d})} + \|\mu_2^\varepsilon\|_{L^2(\mathbb{R}^{2d})} \leq C_\mu, \tag{3.57}$$

and

$$\Big\|\frac{\partial \mu_1^\varepsilon(z)}{\partial z} + \sum_{j=1}^{2} k_j \cdot \nabla_{x_j}\mu_1^\varepsilon(z)\Big\|_{L^2(\mathbb{R}^{2d})} + \Big\|\frac{\partial \mu_2^\varepsilon(z)}{\partial z} + \sum_{j=1}^{2} k_j \cdot \nabla_{x_j}\mu_2^\varepsilon(z)\Big\|_{L^2(\mathbb{R}^{2d})} \leq C_\mu, \tag{3.58}$$

for all $z \in [0, L]$ and $V \in \mathcal{V}$.

The proof of this lemma is very similar to that of Lemma 3.5.1 and is therefore omitted.

Unlike the first moment case, the averaged operator $\mathcal{L}_2^\varepsilon$ still depends on ε. We therefore do not have strong convergence of the $\tilde{P}^\varepsilon$-martingale $G_{\mu_\varepsilon}^{2,\varepsilon}$ to its limit yet. However, the a priori bound on W_ε in L^2 allows us to characterize the limit of $G_{\mu_\varepsilon}^{2,\varepsilon}$ and show strong convergence. This is shown as follows. The first and last terms in (3.56) that are independent of ε give the contribution:

$$\begin{aligned}
\mathcal{L}_2\mu = &\int_0^\infty dr \int_{\mathbb{R}^d} \frac{dp}{(2\pi)^d}\Big[\tilde{R}(r, p-k_1)e^{ir\frac{p^2-k_1^2}{2}}(\mu(z,x_1,p,x_2,k_2) - \mu(z,x_1,k_1,x_2,k_2))\\
&+\tilde{R}(r, k_1-p)e^{ir\frac{k_1^2-p^2}{2}}(\mu(z,x_1,p_1,x_2,k_2) - \mu(z,x_1,k_1,x_2,k_2))\\
&+\tilde{R}(z, p-k_2)e^{ir\frac{p^2-k_2^2}{2}}(\mu(z,x_1,k_1,x_2,p) - \mu(z,x_1,k_1,x_2,k_2))\\
&+\tilde{R}(z, k_2-p)e^{ir\frac{k_2^2-p^2}{2}}(\mu(z,x_1,k_1,x_2,p) - \mu(z,x_1,k_1,x_2,k_2))\Big]\\
= &\int_{\mathbb{R}^d} \frac{dp}{(2\pi)^d}\hat{R}(\frac{p^2-k_1^2}{2}, p-k_1)(\mu(z,x_1,p,x_2,k_2) - \mu(z,x_1,k_1,x_2,k_2))\\
&+\hat{R}(\frac{p^2-k_2^2}{2}, p-k_2)(\mu(z,x_1,k_1,x_2,p) - \mu(z,x_1,k_1,x_2,k_2)).
\end{aligned}$$

The two remaining terms give a contribution that tends to 0 as $\varepsilon \to 0$ for sufficiently smooth test functions. They are given by

$$\begin{aligned}
(\mathcal{L}_2^\varepsilon - \mathcal{L}_2)\mu = &\frac{1}{(2\pi)^d}\int_0^\infty dr \int_{\mathbb{R}^d} dp\tilde{R}(r,p)\Big[\Big(e^{ip\cdot\frac{x_2-x_1}{\varepsilon}+irk_2\cdot p} + e^{irk_1\cdot p+i\frac{x_1-x_2}{\varepsilon}\cdot p}\Big)\\
&\times\Big(\mu(z,x_1,k_1+\frac{p}{2},x_2,k_2+\frac{p}{2}) - \mu(z,x_1,k_1+\frac{p}{2},x_2,k_2-\frac{p}{2})\Big)\\
&+\Big(e^{ip\cdot\frac{x_2-x_1}{\varepsilon}+irk_2\cdot p} + e^{irk_1\cdot p+i\frac{x_1-x_2}{\varepsilon}\cdot p}\Big)\\
&\times\Big(\mu(z,x_1,k_1-\frac{p}{2},x_2,k_2-\frac{p}{2}) - \mu(z,x_1,k_1-\frac{p}{2},x_2,k_2+\frac{p}{2})\Big)\Big].
\end{aligned}$$

We have

$$\tilde{R}(z,p) = \tilde{R}(-z,-p) \geq 0$$

by Bochner's theorem. Since $(\mathcal{L}_2^\varepsilon - \mathcal{L}_2)$ and λ are real quantities, we can take the real part of the above term and, after the change of variables $r \to -r$ and $p \to -p$, obtain

$$\begin{aligned}
(\mathcal{L}_2^\varepsilon - \mathcal{L}_2)\mu &= \frac{1}{(2\pi)^d}\int_{-\infty}^{\infty} dr \int_{\mathbb{R}^d} dp\tilde{R}(r,p)\cos(p\cdot\frac{x_2-x_1}{\varepsilon})(e^{irk_2\cdot p}+e^{irk_1\cdot p}) \\
&\times\Big(\mu(z,x_1,k_1+\frac{p}{2},x_2,k_2+\frac{p}{2})+\mu(z,x_1,k_1-\frac{p}{2},x_2,k_2-\frac{p}{2}) \\
&-\mu(z,x_1,k_1+\frac{p}{2},x_2,k_2-\frac{p}{2})-\mu(z,x_1,k_1-\frac{p}{2},x_2,k_2+\frac{p}{2})\Big) \\
&= \frac{2}{(2\pi)^d}\int_{\mathbb{R}^d} dp(\hat{R}(-k_1\cdot p,p)+\hat{R}(-k_2\cdot p,p))\cos(p\cdot\frac{x_2-x_1}{\varepsilon}) \\
&\times\Big(\mu(z,x_1,k_1+\frac{p}{2},x_2,k_2+\frac{p}{2})-\mu(z,x_1,k_1-\frac{p}{2},x_2,k_2+\frac{p}{2})\Big) \\
&= g_1+g_2+g_3+g_4+c.c.
\end{aligned}$$

We have (since μ is real-valued)

$$\begin{aligned}
I &= \int_{\mathbb{R}^{4d}} dx_1dk_1dx_2dk_2|g_1(z,x_1,k_1,x_2,k_2)|^2 \\
&= C\int_{\mathbb{R}^{6d}} dx_1dk_1dx_2dk_2dpdq\hat{R}(-k_1\cdot p,p)\hat{R}(-k_1\cdot q,q) \\
&\times e^{i(p-q)\cdot\frac{x_2-x_1}{\varepsilon}}\mu(z,x_1,k_1-\frac{p}{2},x_2,k_2+\frac{p}{2})\mu(z,x_1,k_1-\frac{q}{2},x_2,k_2+\frac{q}{2}).
\end{aligned}$$

Using density arguments we may assume that μ has the form

$$\mu(x_1,k_1,x_2,k_2) = \mu_1(x_1-x_2)\mu_2(x_1+x_2)\mu_3(k_1)\mu_4(k_2).$$

Then we have

$$\begin{aligned}
I &= C\int_{\mathbb{R}^{6d}} dx_1dk_1dx_2dk_2dpdq\hat{R}(-k_1\cdot p,p)\hat{R}(-k_1\cdot q,q) \\
&\times e^{-i(p-q)\cdot\frac{x_1}{\varepsilon}}\mu_1^2(x_1)\mu_2^2(x_2)\mu_3(k_1-\frac{p}{2})\mu_4(k_2+\frac{p}{2})\mu_3(k_1-\frac{q}{2})\mu_4(k_2+\frac{q}{2}) \\
&= C\|\mu_2\|_{L^2}^2\int_{\mathbb{R}^{4d}} dk_1dk_2dpdq\hat{R}(-k_1\cdot p,p)\hat{R}(-k_1\cdot q,q)\hat{\nu}(\frac{p-q}{\varepsilon}) \\
&\times\mu_3(k_1-\frac{p}{2})\mu_4(k_2+\frac{p}{2})\mu_3(k_1-\frac{q}{2})\mu_4(k_2+\frac{q}{2})
\end{aligned}$$

where $\nu(x) = \mu_1^2(x)$. We introduce $G(p) = \sup_\omega \hat{R}(\omega,p)$ and use the Cauchy-Schwarz inequality in k_1 and k_2:

$$|I| \le C\|\mu_2\|_{L^2}^2\|\mu_3\|_{L^2}^2\|\mu_4\|_{L^2}^2\int_{\mathbb{R}^{2d}} dpdqG(p)G(q)\left|\hat{\nu}(\frac{p-q}{\varepsilon})\right|.$$

We use again the Cauchy-Schwarz inequality, now in p, to get

$$\begin{aligned}
|I| &\le C\|\mu_2\|_{L^2}^2\|\mu_3\|_{L^2}^2\|\mu_4\|_{L^2}^2\|G\|_{L^2}\int_{\mathbb{R}^d} dqG(q)\left(\int_{\mathbb{R}^d} dp\left|\hat{\nu}(\frac{p}{\varepsilon})\right|^2\right)^{1/2} \\
&\le C\varepsilon^{d/2}\|\mu_2\|_{L^2}^2\|\mu_3\|_{L^2}^2\|\mu_4\|_{L^2}^2\|G\|_{L^2}\|G\|_{L^1}\|\nu\|_{L^2}.
\end{aligned}$$

This proves that $\|(\mathcal{L}_2^\varepsilon - \mathcal{L}_2)\mu\|_{L^2} \to 0$ as $\varepsilon \to 0$. Note that oscillatory integrals of the form

$$\int_{\mathbb{R}^d} e^{i\frac{p\cdot x}{\varepsilon}}\mu(p)dp, \qquad (3.59)$$

are not small in the bigger space $\mathcal{A}'$, which is natural in the context of Wigner transforms. In this bigger space, we cannot control $(\mathcal{L}_2^\varepsilon - \mathcal{L}_2)\mu$ and actually suspect that the limit measure P may no longer be deterministic.

We therefore deduce that

$$G_\mu^2 = \langle W \otimes W, \mu(\hat{V})\rangle(z) - \int_0^z \Big\langle W \otimes W, \Big(\frac{\partial}{\partial z} + k_1 \cdot \nabla_{x_1} + k_2 \cdot \nabla_{x_2} + \mathcal{L}_2\Big)\mu \Big\rangle (s) ds$$

is an approximate $\tilde{P}_\varepsilon$ martingale. The limit of the second moment

$$W_2(z, x_1, k_1, x_2, k_2) = \mathbb{E}^P \{W(z, x_1, k_1) W(z, x_2, k_2)\}$$

thus satisfies (weakly) the transport equation

$$\frac{\partial W_2}{\partial t} + (k_1 \cdot \nabla_{x_1} + k_2 \cdot \nabla_{x_2}) W_2 = \mathcal{L}_2 W_2,$$

with initial data $W_2(0, x_1, k_1, x_2, k_2) = W_0(x_1, k_1) W_0(x_2, k_2)$. Moreover, the operator $\mathcal{L}_2$ acting on a tensor product $\lambda \otimes \lambda$ has the form

$$\mathcal{L}_2[\lambda \otimes \lambda] = \mathcal{L}\lambda \otimes \lambda + \lambda \otimes \mathcal{L}\lambda.$$

This implies that

$$\mathbb{E}^P \{W(z, x_1, k_1) W(z, x_2, k_2)\} = \mathbb{E}^P \{W(z, x_1, k_1)\} \mathbb{E}^P \{W(z, x_2, k_2)\}$$

by uniqueness of the solution to the above transport equation with initial conditions given by $W_0(x_1, k_1) W_0(x_2, k_2)$. This proves that the limiting measure P is deterministic and unique (because characterized by the transport equation) and that the sequence $W_\varepsilon(z, x, k)$ converges in probability to $W(z, x, k)$.

3.5.3 Tightness of P_ε

We now show tightness of the measures P_ε in $\mathcal{X}$. We have the lemma

Lemma 3.5.4 *The family of measures P_ε is weakly compact.*

The proof is as follows; see [12]. A theorem of Mitoma and Fouque [50, 29] implies that in order to verify tightness of the family P_ε it is enough to check that for each $\lambda \in C^1([0, L], \mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d))$ the family of measures $\mathcal{P}_\varepsilon$ on $C([0, L]; \mathbb{R})$ generated by the random processes $W_\lambda^\varepsilon(z) = \langle W_\varepsilon(z), \lambda\rangle$ is tight. Tightness of $\mathcal{P}_\varepsilon$ follows from the following two conditions. First, a Kolmogorov moment condition [11] in the form

$$E^{P_\varepsilon} \{|\langle W, \lambda\rangle(z) - \langle W, \lambda\rangle(z_1)|^\gamma \, |\langle W, \lambda\rangle(z_1) - \langle W, \lambda\rangle(s)|^\gamma\} \leq C_\lambda (z - s)^{1+\beta}, \quad 0 \leq s \leq z \leq L \tag{3.60}$$

should hold with $\gamma > 0$, $\beta > 0$ and C_λ independent of ε. Second, we should have

$$\lim_{R\to\infty} \limsup_{\varepsilon\to 0} \mathrm{Prob}^{\mathcal{P}_\varepsilon} \left\{ \sup_{0\leq z\leq L} |\langle W, \lambda\rangle(z)| > R \right\} = 0.$$

The second condition holds automatically in our case since the process $W_\lambda^\varepsilon(z)$ is uniformly bounded for all $z > 0$ and $\varepsilon > 0$. In order to verify (3.60), note that we have

$$\langle W(z), \lambda\rangle = G_{\lambda_\varepsilon}^\varepsilon(z) - \sqrt{\varepsilon}\langle W, \lambda_1^\varepsilon\rangle - \varepsilon\langle W, \lambda_2^\varepsilon\rangle + \int_0^z ds \langle W, \frac{\partial \lambda}{\partial z} + k \cdot \nabla_x \lambda + \mathcal{L}\lambda\rangle(s) + \sqrt{\varepsilon} \int_0^z ds \langle W, \zeta_\varepsilon^\lambda\rangle(s).$$

The uniform bound (3.47) on $\zeta_\varepsilon^\lambda$ and the bounds on $\|\lambda_{1,2}^\varepsilon(z)\|_{L^2(\mathbb{R}^{2d})}$ in Lemma 3.5.1 imply that it suffices to check (3.60) for

$$x_\varepsilon(z) = G^\varepsilon_{\lambda_\varepsilon}(z) + \int_0^z ds\langle W, \frac{\partial\lambda}{\partial z} + k\cdot\nabla_x\lambda + \mathcal{L}\lambda\rangle(s).$$

We have

$$\begin{aligned}
&E\left\{\left|x_\varepsilon(z)-x_\varepsilon(s)\right|^2\Big|\,\mathcal{F}_s\right\} \le 2E\left\{\left|\int_s^z d\tau\langle W, \frac{\partial\lambda}{\partial z} + k\cdot\nabla_x\lambda + \mathcal{L}\lambda\rangle(\tau)\right|^2\Big|\,\mathcal{F}_s\right\}\\
&+2E\left\{\left|G^\varepsilon_{\lambda_\varepsilon}(z)-G^\varepsilon_{\lambda_\varepsilon}(s)\right|^2\Big|\,\mathcal{F}_s\right\} \le C(z-s)^2 + 2E\left\{\langle G^\varepsilon_{\lambda_\varepsilon}\rangle(z) - \langle G^\varepsilon_{\lambda_\varepsilon}\rangle(s)\big|\,\mathcal{F}_s\right\}.
\end{aligned}$$

Here $\langle G^\varepsilon_{\lambda_\varepsilon}\rangle$ is the increasing process associated with $G^\varepsilon_{\lambda_\varepsilon}$. We will now compute it explicitly. First we obtain that

$$\frac{d}{dh}E^{P_\varepsilon}_{W,\hat V,t}\left\{\langle W,\lambda_\varepsilon\rangle^2(z+h)\right\}\Big|_{h=0} = 2\langle W,\lambda_\varepsilon\rangle\langle W, \frac{\partial\lambda}{\partial z} + k\cdot\nabla_x\lambda_\varepsilon + \frac{1}{\sqrt{\varepsilon}}\mathcal{K}[\hat V,\frac{x}{\varepsilon}]\lambda_\varepsilon\rangle + \frac{1}{\varepsilon}Q\left[\langle W,\lambda_\varepsilon\rangle^2\right]$$

so that

$$\langle W,\lambda_\varepsilon\rangle^2(z) - \int_0^z\left(2\langle W,\lambda_\varepsilon\rangle(s)\langle W, \frac{\partial\lambda}{\partial z} + k\cdot\nabla_x\lambda_\varepsilon + \frac{1}{\sqrt{\varepsilon}}\mathcal{K}[\hat V,\frac{x}{\varepsilon}]\lambda_\varepsilon\rangle(s) + \frac{1}{\varepsilon}Q\left[\langle W,\lambda_\varepsilon\rangle^2\right](s)\right)ds$$

is a martingale. Therefore we have

$$\begin{aligned}
\langle G^\varepsilon_{\lambda_\varepsilon}(z)\rangle &= \int_0^z ds\left[\frac{1}{\varepsilon}Q[\langle W,\lambda_\varepsilon\rangle^2] - \frac{2}{\varepsilon}\langle W,\lambda_\varepsilon\rangle\langle W,Q\lambda_\varepsilon\rangle\right](s)\\
&= \int_0^z ds\left(Q\left[\langle W,\lambda_1^\varepsilon\rangle^2\right] - \langle W,\lambda_1^\varepsilon\rangle\langle W,Q\lambda_1^\varepsilon\rangle(s)\right) + \sqrt{\varepsilon}\int_0^z dsH_\varepsilon(s)
\end{aligned}$$

with

$$\begin{aligned}
H_\varepsilon = {}& 2\sqrt{\varepsilon}\left(Q[\langle W,\lambda_1^\varepsilon\rangle\langle W,\lambda_2^\varepsilon\rangle] - \langle W,\lambda_1^\varepsilon\rangle\langle W,Q\lambda_2^\varepsilon\rangle - \langle W,\lambda_2^\varepsilon\rangle\langle W,Q\lambda_1^\varepsilon\rangle\right)\\
&+\varepsilon\left(Q[\langle W,\lambda_2^\varepsilon\rangle^2] - 2\langle W,\lambda_2^\varepsilon\rangle\langle W,Q\lambda_2^\varepsilon\rangle\right).
\end{aligned}$$

The boundedness of λ_2^ε and that of Q on $L^\infty(\mathcal{V})$ imply that $|H_\varepsilon(s)| \le C$ for all $V\in\mathcal{V}$. This yields

$$E\left\{\langle G^\varepsilon_{\lambda_\varepsilon}\rangle(z) - \langle G^\varepsilon_{\lambda_\varepsilon}\rangle(s)\big|\,\mathcal{F}_s\right\} \le C(z-s)$$

whence

$$E\left\{\left|x_\varepsilon(z)-x_\varepsilon(s)\right|^2\Big|\,\mathcal{F}_s\right\} \le C(z-s).$$

In order to obtain (3.60) we note that

$$\begin{aligned}
&E^{P_\varepsilon}\left\{|x_\varepsilon(z)-x_\varepsilon(z_1)|^\gamma\,|x_\varepsilon(z_1)-x_\varepsilon(s)|^\gamma\right\}\\
&= E^{P_\varepsilon}\left\{E^{P_\varepsilon}\left\{|x_\varepsilon(z)-x_\varepsilon(z_1)|^\gamma\big|\,\mathcal{F}_{z_1}\right\}|x_\varepsilon(z_1)-x_\varepsilon(s)|^\gamma\right\}\\
&\le E^{P_\varepsilon}\left\{\left[E^{P_\varepsilon}\left\{|x_\varepsilon(z)-x_\varepsilon(z_1)|^2\Big|\,\mathcal{F}_{z_1}\right\}\right]^{\gamma/2}|x_\varepsilon(z_1)-x_\varepsilon(s)|^\gamma\right\}\\
&\le C(z-z_1)^{\gamma/2}E^{P_\varepsilon}\left\{|x_\varepsilon(z_1)-x_\varepsilon(s)|^\gamma\right\} \le C(z-z_1)^{\gamma/2}E^{P_\varepsilon}\left\{E^{P_\varepsilon}\left\{|x_\varepsilon(z_1)-x_\varepsilon(s)|^\gamma\,|\mathcal{F}_s\right\}\right\}\\
&\le C(z-z_1)^{\gamma/2}E^{P_\varepsilon}\left\{\left[E^{P_\varepsilon}\left\{|x_\varepsilon(z_1)-x_\varepsilon(s)|^2\Big|\,\mathcal{F}_s\right\}\right]^{\gamma/2}\right\} \le C(z-z_1)^{\gamma/2}(z_1-s)^{\gamma/2}\\
&\le C(z-s)^\gamma.
\end{aligned}$$

Choosing now $\gamma > 1$ we get (3.60) which finishes the proof of Lemma 3.5.4.

3.5.4 Remarks

Statistical stability and a priori bounds. As we have already mentioned, the uniform L^2 bound for the Wigner transform is crucial in the derivation of Thm. 3.4.1. In the absence of an a priori L^2 bound, we are not able to characterize the limiting measure P. However we can characterize its first moment. The derivation is done in [2]. Let us assume that W_ε is bounded in $\mathcal{A}'$, the space of distributions dual to

$$\|f\|_{\mathcal{A}} = \int \sup_x |\tilde{f}(x,y)| dy,$$

as is the case for the Wigner transform of a pure state ψ_ε uniformly bounded in $L^2(\mathbb{R}^d)$. Then we can show that $\mathbb{E}^{P_\varepsilon}\{W_\varepsilon\}$ converges weakly to $\overline{W}$, solution of (3.27), with appropriate initial conditions (the Wigner transform of the limit $\psi_\varepsilon(0,x)$). The proof is very similar to that obtained above, except that in the proof of convergence, as well as in the proof of tightness of the sequence of measures P_ε (now defined on a ball in $\mathcal{C}([0,L];\mathcal{A}')$), we need to show that the test functions $\lambda_{1,2}$ are bounded in $\mathcal{A}'$ rather than $L^2(\mathbb{R}^{2d})$.

However the proof of convergence of the second martingale in section 3.5.2 *does not* extend to the case of a uniform bound in $\mathcal{A}'$. Technically, the obstacle resides in the fact that the oscillatory integrals (3.59) are small in $L^2(\mathbb{R}^{2d})$ but *not* in $\mathcal{A}'$. Since $\mathcal{A}'$ includes bounded measures, any measure $\mu(dp)$ concentrating on the hyperplane orthogonal to x will render the integral (3.59) an order $O(1)$ quantity.

The above discussion does not provide proof that P_ε does not converge to a deterministic limit. However it strongly suggests that if W_ε is allowed to become quite singular in $\mathcal{A}'$, then on these paths P_ε may not become sufficiently self-averaging to converge to a deterministic limit. Actually, in the simplified regime of the Itô-Schrödinger equation (a further simplification compared to the paraxial wave equation), it is shown in [3] that the measure P_ε does not converge to a limiting deterministic measure when the initial Wigner measure is very singular (converges to a delta function in both x and k). Instead, *scintillation* effects, which measure the distance between the second moment of W_ε and the square of its first moment, are shown to persist for all finite times (for an appropriate scaling). This does not characterize the limiting measure P either (this remains an open problem even in the Itô-Schrödinger framework), but at least shows that P is not deterministic.

Paraxial and radiative transfer regimes. Note that in the limit where the potential $V(z,x)$ oscillates very slowly in the z variable, so that $R(z,x)$ converges to a function that does not depend on z (because $V(z,x)$ becomes highly correlated in z), whence $\hat{R}(\omega,p)$ converges to a function of the form $\delta(\omega)\hat{R}(p)$, we obtain the limiting average transport equation

$$\kappa\frac{\partial\overline{W}}{\partial z} + k\cdot\nabla_x\overline{W} = \kappa^2\int_{\mathbb{R}^d}\hat{R}(p-k)\delta\Big(\frac{|k|^2}{2}-\frac{|p|^2}{2}\Big)\Big(W(x,p)-W(x,k)\Big)\frac{dp}{(2\pi)^d}. \qquad (3.61)$$

This is the radiative transfer equation for the Schrödinger equation (3.13) when the potential $V(x)$ is independent of the variable z. We do not recover the full radiative transfer since we started with the paraxial approximation. However we recover the correct radiative transfer equation for the Schrödinger equation.

Note that the dispersion relation for wave equations $\omega = c_0|k|$ is now replaced by its "paraxial" approximation $\omega = |k|^2/2$, where k now is the transverse component of the wavevector only.

Chapter 4

Applications to Time-Reversal Experiments

This chapter is based on [4, 5] but the concept of one-step time reversal emerged during early discussions with Knut Solna.

4.1 Time-reversal experiments

In time reversal experiments, acoustic waves are emitted from a localized source, recorded in time by an array of receivers-transducers, time reversed, and re-transmitted into the medium, so that the signals recorded first are re-emitted last and vice versa [20, 21, 28, 35, 42, 46]: a schematic description of the time reversal procedure is depicted in Fig. 4.1.

Figure 4.1: The Time Reversal Procedure. Top: Propagation of signal and measurements in time. Bottom: Time reversal of recorded signals and back-propagation into the medium.

Early experiments in time reversal acoustics are described in [20]; see also the more recent papers [26, 27, 28] – this list is by no means exhaustive and the literature on the subject is by now vast. The re-transmitted signal refocuses at the location of the original source with a modified shape that depends on the array of receivers. The salient feature of these time reversal experiments is that refocusing is much better when wave propagation occurs in complicated environments than in homogeneous media. Time reversal techniques with improved refocusing in heterogeneous medium have found important applications in medicine, non-destructive testing, underwater acoustics, and wireless communications (see the above references). It has been also applied to imaging in weakly random media [10, 14, 28] and led to a recent concept of coherent interferometric imaging (CINT) of Borcea, Papanicolaou and Tsogka [15, 16, 17].

A very qualitative explanation for the better refocusing observed in heterogeneous media is based on *multipathing*. Since waves can scatter off a larger number of heterogeneities, more paths coming from the source reach the recording array, thus more is known about the source by the transducers than in a homogeneous medium. The heterogeneous medium plays the role of a lens that widens the aperture through which the array of receivers sees the source. Refocusing is also qualitatively justified by ray theory (geometrical optics). The phase shift caused by multiple scattering is exactly compensated when the time reversed signal follows the same path back to the source location. This phase cancellation happens only at the source location. The phase shift along paths leading to other points in space is essentially random. The interference of multiple paths will thus be constructive at the source location and destructive anywhere else. This explains why refocusing at the source location is improved when the number of scatterers is large.

As convincing as they are, the above explanations remain qualitative and do not allow us to quantify how the refocused signal is modified by the time reversal procedure. Quantitative justifications require to analyze wave propagation more carefully. The first quantitative description of time reversal was obtained in [18] in the framework of one-dimensional random media. That paper provides the first mathematical explanation of two of the most prominent features of time reversal: heterogeneities improve refocusing and refocusing occurs for almost every realization of the random medium. Various extensions and generalizations to the three-dimensional layered case, including nonlinear effects, have been done in the work by Garnier, Fouque, Nachbin, Papanicolaou and Solna, and are described in detail in the recent excellent book [30]. The first multi-dimensional quantitative description of time reversal was obtained in [13] for the parabolic approximation, i.e., for waves that propagate in a privileged direction with no backscattering (see also [54] for further analysis of time reversal in this regime). That paper shows that the random medium indeed plays the role of a lens. The back-propagated signal behaves as if the initial array were replaced by another one with a much bigger effective aperture. In a slightly different context, time reversal in ergodic cavities was analyzed in [8]. There, wave mixing is created by reflection at the boundary of a chaotic cavity, which plays a similar role to the heterogeneities in a heterogeneous medium.

In this chapter we consider the theory of time-reversal experiments for general classical waves propagating in weakly fluctuating random media. It is convenient to understand refocusing in time reversal experiments in the following three-step general framework:

(i) A signal propagating from a localized source is recorded at a single time $T > 0$ by an array of receivers.

(ii) The recorded signal is processed at the array location.

(iii) The processed signal is emitted from the array and propagates in the *same* medium during the same amount of time T.

As we will see, this formulation allows us to reduce the mathematical problem of the description of the refocused signal to the question of the passage from the wave equations to the kinetic models. While the latter problem is also difficult, we may apply whatever is known in that area to the time-reversal problems. Accordingly, the mathematical rigor of our statements on time-reversal experiments below depends on the regime of consideration – for instance, they are mostly formal in the radiative transfer regime but are rigorous in the random geometric optics regime (see [6] for the precise statements). To keep the presentation uniform we will concentrate here solely on the transport regime.

The first main result of this chapter is that the repropagated signal will refocus at the location of the original source for a large class of waves and a large class of processings. The experiments described above correspond to the specific processing of acoustic waves in which pressure is kept unchanged and the sign of the velocity field is reversed.

The second main result is a quantitative description of the re-transmitted signal. We show that the re-propagated signal $\mathbf{u}^B(\xi)$ at a point ξ near the source location can be written in the high frequency limit as the following convolution of the original source $\mathbf{S}$

$$\mathbf{u}^B(\xi) = (F * \mathbf{S})(\xi). \tag{4.1}$$

The kernel F depends on the location of the recording array and on the signal processing. The quality of the refocusing depends on the spatial decay of F. It turns out that it can be expressed in terms of the Wigner transform [53] of two wave fields. The decay properties of F depend on the smoothness of the Wigner transform in the phase space and it is here that the kinetic theories becomes useful. Here we consider the high frequency regime when the wavelength of the initial signal is small compared to the distance of propagation. In addition we assume that the wavelength is comparable to the correlation length of the medium. This is the radiative transport regime. It has been extensively studied mathematically for the Schrödinger equation [23, 56] and formally using perturbation expansions for the classical waves [7, 53]. In this regime the Wigner transform satisfies a radiative transport equation, which is used to describe the evolution of the energy density of waves in random media [36, 53, 55, 56]. The transport equations possess a smoothing effect so that the Wigner distribution becomes less singular in random media, which implies a stronger decay of the convolution kernel F and a better refocusing. The diffusion approximation to the radiative transport equations provides simple reconstruction formulas that can be used to quantify the refocusing quality of the back-propagated signal. This construction applies to a large class of classical waves: acoustic, electromagnetic, elastic, and others, and allows for a large class of signal processings at the recording array.

4.2 Classical Time Reversal and One-Step Time Reversal

Propagation of acoustic waves is described by a system of equations for the pressure $p(t,x)$ and acoustic velocity $\mathbf{v}(t,x)$:

$$\begin{aligned} \rho(x)\frac{\partial \mathbf{v}}{\partial t} + \nabla p &= 0 \\ \kappa(x)\frac{\partial p}{\partial t} + \nabla\cdot\mathbf{v} &= 0, \end{aligned} \tag{4.2}$$

with suitable initial conditions and where $\rho(x)$ and $\kappa(x)$ are density and compressibility of the underlying medium, respectively. These equations can be recast as the following linear hyperbolic system

$$A(x)\frac{\partial \mathbf{u}}{\partial t} + D^j\frac{\partial \mathbf{u}}{\partial x^j} = 0, \quad x \in \mathbb{R}^3 \tag{4.3}$$

with the vector $\mathbf{u} = (\mathbf{v}, p) \in \mathbb{C}^4$. The matrix $A = \mathrm{Diag}(\rho,\rho,\rho,\kappa)$ is positive definite. The 4×4 matrices D^j, $j = 1,2,3$, are symmetric and given by $D^j_{mn} = \delta_{m4}\delta_{nj} + \delta_{n4}\delta_{mj}$. We use the Einstein convention of summation over repeated indices.

The time reversal experiments in [20] consist of two steps. First, the direct problem

$$\begin{aligned} &A(x)\frac{\partial \mathbf{u}}{\partial t} + D^j\frac{\partial \mathbf{u}}{\partial x^j} = 0, \ 0 \le t \le T \\ &\mathbf{u}(0,x) = \mathbf{S}(x) \end{aligned} \tag{4.4}$$

with a localized source $\mathbf{S}$ centered at a point x_0 is solved. The signal is recorded during the period of time $0 \le t \le T$ by an array of receivers located at $\Omega \subset \mathbb{R}^3$. Second, the signal is time reversed and re-emitted into the medium. Time reversal is described by multiplying $\mathbf{u} = (\mathbf{v}, p)$ by the matrix $\Gamma = \mathrm{Diag}(-1,-1,-1,1)$. The back-propagated signal solves

$$\begin{aligned} &\frac{\partial \mathbf{u}}{\partial t} + A^{-1}(x)D^j\frac{\partial \mathbf{u}}{\partial x^j} = \frac{1}{T}\mathbf{R}(2T-t,x), \ T \le t \le 2T \\ &\mathbf{u}(T,x) = 0 \end{aligned} \tag{4.5}$$

with the source term

$$\mathbf{R}(t, x) = \Gamma \mathbf{u}(t, x) \chi(x). \tag{4.6}$$

The function $\chi(x)$ is either the characteristic function of the set where the recording array is located, or some other function that allows for possibly space-dependent amplification of the re-transmitted signal.

The back-propagated signal is then given by $\mathbf{u}(2T, x)$. We can decompose it as

$$\mathbf{u}(2T, x) = \frac{1}{T} \int_0^T ds\ \mathbf{w}(s, x; s), \tag{4.7}$$

where the vector-valued function $\mathbf{w}(t, x; s)$ solves the initial value problem

$$\begin{aligned} &A(x) \frac{\partial \mathbf{w}(t, x; s)}{\partial t} + D^j \frac{\partial \mathbf{w}(t, x; s)}{\partial x^j} = 0, \quad 0 \leq t \leq s \\ &\mathbf{w}(0, x; s) = \mathbf{R}(s, x). \end{aligned}$$

We deduce from (4.7) that it is sufficient to analyze the refocusing properties of $\mathbf{w}(s, x; s)$ for $0 \leq s \leq T$ to obtain those of $\mathbf{u}(2T, x)$. For a fixed value of s, we call the construction of $\mathbf{w}(s, x; s)$ one-step time reversal.

We define one-step time reversal more generally as follows. The direct problem (4.4) is solved until time $t = T$ to yield $\mathbf{u}(T^-, x)$. At time T, the signal is recorded and processed. The processing is modeled by an amplification function $\chi(x)$, a blurring kernel $f(x)$, and a (possibly spatially varying) time reversal matrix Γ. After processing, we have

$$\mathbf{u}(T^+, x) = \Gamma(f * (\chi \mathbf{u}))(T^-, x) \chi(x). \tag{4.8}$$

The processed signal then propagates for the same amount of time T:

$$\begin{aligned} &A(x) \frac{\partial \mathbf{u}}{\partial t} + D^j \frac{\partial \mathbf{u}}{\partial x^j} = 0, \ T \leq t \leq 2T \\ &\mathbf{u}(T^+, x) = \Gamma(f * (\chi \mathbf{u}))(T^-, x) \chi(x). \end{aligned} \tag{4.9}$$

The main question is whether $\mathbf{u}(2T, x)$ refocuses at the location of the original source $\mathbf{S}(x)$ and how the original signal has been modified by the time reversal procedure. Notice that in the case of full ($\Omega = \mathbb{R}^3$) and exact ($f(x) = \delta(x)$) measurements with $\Gamma = \text{Diag}(-1, -1, -1, 1)$, the time-reversibility of first-order hyperbolic systems implies that $\mathbf{u}(2T, x) = \Gamma \mathbf{S}(x)$, which corresponds to exact refocusing. When only partial measurements are available we shall see in the following sections that $\mathbf{u}(2T, x)$ is closer to $\Gamma \mathbf{S}(x)$ when propagation occurs in a heterogeneous medium than in a homogeneous medium.

The pressure field $p(t, x)$ satisfies the following scalar wave equation

$$\frac{\partial^2 p}{\partial t^2} - \frac{1}{\kappa(x)} \nabla \cdot \left(\frac{1}{\rho(x)} \nabla p \right) = 0. \tag{4.10}$$

A schematic description of the one-step procedure for the wave equation is presented in Fig. 4.2. A numerical experiment for the one-step time reversal procedure is shown in Fig. 4.3. In the

Figure 4.2: The One-Step Time Reversal Procedure. Here, p_t denotes $\dfrac{\partial p}{\partial t}$.

numerical simulations, there is no blurring, $f(x) = \delta(x)$, and the array of receivers is the

Figure 4.3: Numerical experiment using the one-step time reversal procedure. Top Left: initial condition $p(0,x)$, a peaked Gaussian of maximal amplitude equal to 1. Top Right: forward solution $p(T^-,x)$, of maximal amplitude 0.04. Bottom Right: recorded solution $p(T^+,x)$, of maximal amplitude 0.015 on the domain $\Omega = (-1/6, 1/6)^2$. Bottom Left: back-propagated solution $p(2T,x)$, of maximal amplitude 0.07.

domain $\Omega = (-1/6, 1/6)^2$ ($\chi(x)$ is the characteristic function of Ω). Note that the truncated signal does not retain any information about the ballistic part of the original wave (the part that propagates without scattering with the underlying medium). In a homogeneous medium, the truncated signal would then be nearly identically zero (not quite zero since the numerics are done in two dimensions) and no refocusing would be observed. The interesting aspect of time reversal is that a coherent signal emerges at time $2T$ out of a signal at time T^+ that seems to have no useful information.

4.3 Theory of Time Reversal in Random Media

Our objective is now to present a theory that explains in a quantitative manner the refocusing properties described in the preceding sections. We consider here the one-step time reversal for acoustic wave. Generalizations to other types of waves and more general processings in (4.9) are given in Section 4.4.

4.3.1 Refocused Signal

We recall that the one-step time reversal procedure consists of letting an initial pulse $\mathbf{S}(x)$ propagate according to (4.4) until time T,

$$\mathbf{u}(T^-,x) = \int_{\mathbb{R}^3} G(T,x;z)\mathbf{S}(z)dz,$$

where $G(T,x;z)$ is the Green's matrix solution of

$$\begin{aligned} &A(x)\frac{\partial G(t,x;y)}{\partial t} + D^j \frac{\partial G(t,x;y)}{\partial x^j} = 0, \; 0 \le t \le T \\ &G(0,x;y) = I\delta(x-y). \end{aligned} \tag{4.11}$$

At time T, the "intelligent" array reverses the signal. For acoustic pulses, this means keeping pressure unchanged and reversing the sign of the velocity field. The array of receivers is located in $\Omega \subset \mathbb{R}^3$. The amplification function $\chi(x)$ is an arbitrary bounded function supported in Ω, such as its characteristic function ($\chi(x) = 1$ for $x \in \Omega$ and $\chi(x) = 0$ otherwise) when all transducers have the same amplification factor. We also allow for some blurring of the recorded data modeled by a convolution with a function $f(x)$. The case $f(x) = \delta(x)$ corresponds to exact measurements. Finally, the signal is time reversed, that is, the direction of the acoustic velocity is reversed. Here, the operator Γ in (4.8) is simply multiplication by the matrix

$$\Gamma = \mathrm{Diag}(-1,-1,-1,1). \tag{4.12}$$

The signal at time T^+ after time reversal takes then the form

$$\mathbf{u}(T^+,x) = \int_{\mathbb{R}^6} \Gamma G(T,y';z)\chi(x)\chi(y')f(x-y')\mathbf{S}(z)dzdy'. \tag{4.13}$$

The last step (4.9) consists of letting the time reversed field propagate through the random medium until time $2T$. To compare this signal with the initial pulse $\mathbf{S}$, we need to reverse the acoustic velocity once again, and define

$$\mathbf{u}^B(x) = \Gamma\mathbf{u}(2T,x) = \int_{\mathbb{R}^9} \Gamma G(T,x;y)\Gamma G(T,y';z)\chi(y)\chi(y')f(y-y')\mathbf{S}(z)dydy'dz. \qquad (4.14)$$

The time reversibility of first-order hyperbolic systems implies that $\mathbf{u}^B(x) = \mathbf{S}(x)$ when $\Omega = \mathbb{R}^d$, $\chi \equiv 1$, and $f(x) = \delta(x)$, that is, when full and non-distorted measurements are available. It remains to understand which features of $\mathbf{S}$ are retained by $\mathbf{u}^B(x)$ when only partial measurement is available.

4.3.2 Localized Source and Scaling

We consider an asymptotic solution of the time reversal problem (4.4), (4.9) when the support λ of the initial pulse $\mathbf{S}(x)$ is much smaller than the distance L of propagation between the source and the recording array: $\varepsilon = \lambda/L \ll 1$. We also take the size a of the array comparable to L: $a/L = O(1)$. We assume that the time T between the emission of the original signal and recording is of order L/c_0, where c_0 is a typical speed of propagation of the acoustic wave. We consequently consider the initial pulse to be of the form

$$\mathbf{u}(0,x) = \mathbf{S}(\frac{x-x_0}{\varepsilon})$$

in non-dimensionalized variables $x' = x/L$ and $t' = t/(L/c_0)$. We drop primes to simplify notation. Here x_0 is the location of the source. The transducers obviously have to be capable of capturing signals of frequency ε^{-1} and blurring should happen on the scale of the source, so we replace $f(x)$ by $\varepsilon^{-d}f(\varepsilon^{-1}x)$. Finally, we are interested in the refocusing properties of $\mathbf{u}^B(x)$ in the vicinity of x_0. We therefore introduce the scaling $x = x_0 + \varepsilon\xi$. With these changes of variables, expression (4.14) is recast as

$$\begin{aligned}\mathbf{u}^B(\xi;x_0) &= \Gamma\mathbf{u}(2T, x_0+\varepsilon\xi) \\ &= \int_{\mathbb{R}^9} \Gamma G(T,x_0+\varepsilon\xi;y)\Gamma G(T,y';x_0+\varepsilon z)\chi(y,y')\mathbf{S}(z)dydy'dz,\end{aligned} \qquad (4.15)$$

where

$$\chi(y,y') = \chi(y)\chi(y')f(\frac{y-y'}{\varepsilon}). \qquad (4.16)$$

In the sequel we will also allow the medium to vary on a scale comparable to the source scale ε. Thus the Green's function G and the matrix A depend on ε. We do not make this dependence explicit to simplify notation. We are interested in the limit of $\mathbf{u}^B(\xi;x_0)$ as $\varepsilon \to 0$.

4.3.3 Adjoint Green's Function

The analysis of the re-propagated signal relies on the study of the two point correlation at nearby points of the Green's matrix in (4.15). There are two undesirable features in (4.15). First, the two nearby points $x_0+\varepsilon\xi$ and $x_0+\varepsilon z$ are terminal and initial points in their respective Green's matrices. Second, one would like the matrix Γ between the two Green's matrices to be outside of their product. However, Γ and G do not commute. For these reasons, we introduce the *adjoint* Green's matrix, solution of

$$\begin{aligned}&\frac{\partial G_*(t,x;y)}{\partial t}A(x) + \frac{\partial G_*(t,x;y)}{\partial x^j}D^j = 0 \\ &G_*(0,x;y) = A^{-1}(x)\delta(x-y).\end{aligned} \qquad (4.17)$$

We now prove that

$$G_*(t,x;y) = \Gamma G(t,y;x)A^{-1}(x)\Gamma. \tag{4.18}$$

Note that for all initial data $\mathbf{S}(x)$, the solution $\mathbf{u}(t,x)$ of (4.4) satisfies

$$\mathbf{u}(t,x) = \int_{\mathbb{R}^d} G(t-s,x;y)\mathbf{u}(s,y)dy$$

for all $0 \le s \le t \le T$ since the coefficients in (4.4) are time-independent. Differentiating the above with respect to s and using (4.4) yields

$$0 = \int_{\mathbb{R}^d} \Big(- \frac{\partial G(t-s,x;y)}{\partial t}\mathbf{u}(s,y) - G(t-s,x;y)A^{-1}(y)D^j \frac{\partial \mathbf{u}(s,y)}{\partial y^j}\Big) dy$$

Upon integrating by parts and letting $s = 0$, we get

$$0 = \int_{\mathbb{R}^d} \Big(- \frac{\partial G(t,x;y)}{\partial t} + \frac{\partial}{\partial y^j}\left[G(t,x;y)A^{-1}(y)D^j\right]\Big)\mathbf{S}(y)dy.$$

Since the above relation holds for all test functions $\mathbf{S}(y)$, we deduce that

$$\frac{\partial G(t,x;y)}{\partial t} - \frac{\partial}{\partial y^j}\left[G(t,x;y)A^{-1}(y)D^j\right] = 0. \tag{4.19}$$

Interchanging x and y in the above equation and multiplying it on the left and the right by Γ, we obtain that

$$\frac{\partial}{\partial t}\left[\Gamma G(t,y;x)A^{-1}(x)\right]A(x)\Gamma - \frac{\partial}{\partial x^j}\left[\Gamma G(t,y;x)A^{-1}(x)\right]D^j\Gamma = 0. \tag{4.20}$$

We remark that

$$\Gamma D^j = -D^j\Gamma \qquad \text{and} \qquad \Gamma A(x) = A(x)\Gamma, \tag{4.21}$$

so that

$$\frac{\partial}{\partial t}\left[\Gamma G(t,y;x)A^{-1}(x)\Gamma\right]A(x) + \frac{\partial}{\partial x^j}\left[\Gamma G(t,y;x)A^{-1}(x)\Gamma\right]D^j = 0$$

with $\Gamma G(0,y;x)A^{-1}(x)\Gamma = A^{-1}(x)\delta(x-y)$. Thus (4.18) follows from the uniqueness of the solution to the above hyperbolic system with given initial conditions. We can now recast (4.15) as

$$\begin{aligned}\mathbf{u}^B(\xi;x_0) \;=\; &\int_{\mathbb{R}^9} \Gamma G(T,x_0+\varepsilon\xi;y)G_*(T,x_0+\varepsilon z;y')\Gamma \\ &\qquad \times\chi(y)\chi(y')f(\frac{y-y'}{\varepsilon})A(x_0+\varepsilon z)\mathbf{S}(z)dydy'dz.\end{aligned} \tag{4.22}$$

One may further simplify (4.22) with the help of the auxiliary matrix-valued functions $Q(t,x;q)$ and $Q_*(t,x,q)$ defined by

$$\begin{aligned} Q(T,x;q) &= \int_{\mathbb{R}^d} G(T,x;y)\chi(y)e^{iq\cdot y/\varepsilon}dy, \\ Q_*(T,x;q) &= \int_{\mathbb{R}^3} G_*(T,x;y)\chi(y)e^{-iq\cdot y/\varepsilon}dy. \end{aligned} \tag{4.23}$$

They solve the hyperbolic systems of equations (4.4) and (4.17) with initial conditions given by $Q(0,x;q) = \chi(x)e^{iq\cdot x/\varepsilon}I$ and $Q_*(0,x;q) = A^{-1}(x)\chi(x)e^{-iq\cdot x/\varepsilon}$, respectively. Thus (4.22) becomes

$$\mathbf{u}^B(\xi;x_0) = \int_{\mathbb{R}^6}\Gamma Q(T,x_0+\varepsilon\xi;q)Q_*(T,x_0+\varepsilon z;q)\Gamma A(x_0+\varepsilon z)\mathbf{S}(z)\hat{f}(q)\frac{dqdz}{(2\pi)^3}, \tag{4.24}$$

where $\hat{f}(q) = \int_{\mathbb{R}^d} e^{-iq\cdot x}f(x)dx$ is the Fourier transform of $f(x)$.

4.3.4 Wigner Transform

The back-propagated signal in (4.24) now has the suitable form to be analyzed in the Wigner transform formalism [31, 53]. We define

$$W_\varepsilon(t,x,k) = \int_{\mathbb{R}^d} \hat{f}(q) U_\varepsilon(t,x,k;q)dq, \tag{4.25}$$

where

$$U_\varepsilon(t,x,k;q) = \int_{\mathbb{R}^d} e^{ik\cdot y} Q(t, x - \frac{\varepsilon y}{2}; q) Q_*(t, x + \frac{\varepsilon y}{2}; q) \frac{dy}{(2\pi)^3}. \tag{4.26}$$

Taking the inverse Fourier transform we verify that

$$Q(t,x;q)Q_*(t,y;q) = \int_{\mathbb{R}^3} e^{-ik\cdot(y-x)/\varepsilon} U_\varepsilon(t, \frac{x+y}{2}, k; q)dk,$$

hence

$$\mathbf{u}^B(\xi;x_0) = \int_{\mathbb{R}^6} e^{ik\cdot(\xi - z)} \Gamma W_\varepsilon(T, x_0 + \varepsilon\frac{z+\xi}{2}, k) \Gamma A(x_0 + \varepsilon z) \mathbf{S}(z) \frac{dzdk}{(2\pi)^3}. \tag{4.27}$$

We have thus reduced the analysis of $\mathbf{u}(\xi;x_0)$ as $\varepsilon \to 0$ to that of the asymptotic properties of the Wigner transform W_ε. The Wigner transform has been used extensively in the study of wave propagation in random media, especially in the derivation of radiative transport equations modeling the propagation of high frequency waves. We refer to [31, 49, 53]. Note that in the usual definition of the Wigner transform, one has the adjoint matrix Q^* in place of Q_* in (4.26). This difference is not essential since Q_* and Q^* satisfy the same evolution equation, though with different initial data.

The main reason for using the Wigner transform in (4.27) is that W_ε has a weak limit W as $\varepsilon \to 0$. Its existence follows from simple a priori bounds for $W_\varepsilon(t,x,k)$. Let us introduce the space $\mathcal{A}$ of matrix-valued functions $\phi(x,k)$ bounded in the norm $\|\cdot\|_{\mathcal{A}}$ defined by

$$\|\phi\|_{\mathcal{A}} = \int_{\mathbb{R}^3} \sup_x \|\tilde{\phi}(x,y)\| dy, \qquad \text{where} \quad \tilde{\phi}(x,y) = \int_{\mathbb{R}^3} e^{-ik\cdot y}\phi(x,k)dk.$$

We denote by $\mathcal{A}'$ its dual space, which is a space of distributions large enough to contain matrix-valued bounded measures, for instance. We then have the following result:

Lemma 4.3.1 *Let $\chi(x) \in L^2(\mathbb{R}^3)$ and $\hat{f}(q) \in L^1(\mathbb{R}^3)$. Then there is a constant $C > 0$ independent of $\varepsilon > 0$ and $t \in [0,\infty)$ such that for all $t \in [0,\infty)$, we have $\|W_\varepsilon(t,x,k)\|_{\mathcal{A}'} < C$.*

The proof of this lemma is essentially contained in [31, 49], see also [4]. One may actually get L^2-bounds for W_ε in our setting because of the regularizing effect of $\hat{f}$ in (4.25) but this is not essential for the purposes of this chapter as we are working on a formal level. However, this setting is one example when the mixture of states arises naturally. This is also crucial for trhe rigorous justification of the analog of the results of this chapter in the geometic optics regime in [6].

We therefore obtain the existence of a subsequence $\varepsilon_k \to 0$ such that W_{ε_k} converges weakly to a distribution $W \in \mathcal{A}'$. Moreover, an easy calculation shows that at time $t = 0$, we have

$$W(0,x_0,k) = |\chi(x_0)|^2 A_0^{-1}(x_0)\hat{f}(k). \tag{4.28}$$

Here, $A_0 = A$ when A is independent of ε, and $A_0 = \lim_{\varepsilon\to 0} A_\varepsilon$ if we assume that the family of matrices $A_\varepsilon(x)$ is uniformly bounded and continuous with the limit A_0 in $\mathcal{C}(\mathbb{R}^d)$. These assumptions on A_ε are sufficient to deal with the radiative transport regime we will consider in section 4.3.7. Under the same assumptions on A_ε, we have the following result.

Proposition 4.3.2 *The back-propagated signal* $\mathbf{u}^B(\xi;x_0)$ *given by (4.27) converges weakly in* $\mathcal{S}'(\mathbb{R}^3\times\mathbb{R}^3)$ *as* $\varepsilon\to 0$ *to the limit*

$$\mathbf{u}^B(\xi;x_0)=\int_{\mathbb{R}^6} e^{ik\cdot(\xi-z)}\Gamma W(T,x_0,k)\Gamma A_0(x_0)\mathbf{S}(z)\frac{dzdk}{(2\pi)^3}. \tag{4.29}$$

The proof of this proposition is based on taking the duality product of $\mathbf{u}^B(\xi;x_0)$ with a vector-valued test function $\boldsymbol{\phi}(\xi;x_0)$ in $\mathcal{S}(\mathbb{R}^3\times\mathbb{R}^3)$. After a change of variables we obtain $\langle \mathbf{u}^B,\boldsymbol{\phi}\rangle = \langle W_\varepsilon, Z_\varepsilon\rangle$. Here the duality product for matrices is given by the trace $\langle A,B\rangle=\sum_{i,k}\langle A_{ik},B_{ik}\rangle$, and

$$Z_\varepsilon(x_0,k)=\int_{\mathbb{R}^6} e^{ik\cdot(z-\xi)}\Gamma\boldsymbol{\phi}(\xi,x_0-\varepsilon\frac{z+\xi}{2})\mathbf{S}^*(z)A_\varepsilon(x_0+\varepsilon\frac{z-\xi}{2})\Gamma\frac{dzd\xi}{(2\pi)^3}. \tag{4.30}$$

Defining Z as the limit of Z_ε as $\varepsilon\to 0$ by replacing formally ε by 0 in the above expression, (4.29) follows from showing that $\|Z_\varepsilon - Z\|_{\mathcal{A}}\to 0$ as $\varepsilon\to 0$. This is straightforward and we omit the details.

The above proposition tells us how to reconstruct the back-propagated solution in the high frequency limit from the limit Wigner matrix W. Notice that we have made almost no assumptions on the medium described by the matrix $A_\varepsilon(x)$. At this level, the medium can be either homogeneous or heterogeneous, and the particular scale of oscillations is not important as long as $A_\varepsilon(x)$ strongly converge to A_0. Without any further assumptions, we can also obtain some information about the matrix W. Let us define the dispersion matrix for the system (4.4) as [53]

$$L(x,k)=A_0^{-1}(x)k_jD^j. \tag{4.31}$$

It is given explicitly by

$$L(x,k)=\begin{pmatrix} 0 & 0 & 0 & k_1/\rho(x)\\ 0 & 0 & 0 & k_2/\rho(x)\\ 0 & 0 & 0 & k_3/\rho(x)\\ k_1/\kappa(x) & k_2/\kappa(x) & k_3/\kappa(x) & 0\end{pmatrix}.$$

The matrix L has a double eigenvalue $\omega_0=0$ and two simple eigenvalues $\omega_\pm(x,k)=\pm c(x)|k|$, where $c(x)=1/\sqrt{\rho(x)\kappa(x)}$ is the speed of sound. The eigenvalues $\omega_\pm$ are associated with eigenvectors $\mathbf{b}_\pm(x,k)$ and the eigenvalue $\omega_0=0$ is associated with the eigenvectors $\mathbf{b}_j(x,k)$, $j=1,2$. They are given by

$$\mathbf{b}_\pm(x,k)=\begin{pmatrix} \pm\dfrac{\hat{k}}{\sqrt{2\rho(x)}}\\ \dfrac{1}{\sqrt{2\kappa(x)}}\end{pmatrix},\qquad \mathbf{b}_j(x,k)=\begin{pmatrix}\dfrac{\mathbf{z}^j(k)}{\sqrt{\rho(x)}}\\ 0\end{pmatrix}, \tag{4.32}$$

where $\hat{k}=k/|k|$ and $z^1(k)$ and $z^2(k)$ are chosen so that the triple $(\hat{k},z^1(k),z^2(k))$ forms an orthonormal basis. The eigenvectors are normalized so that

$$(A_0(x)\mathbf{b}_j(x,k)\cdot\mathbf{b}_k(x,k))=\delta_{jk}, \tag{4.33}$$

for all $j,k\in J=\{+,-,1,2\}$. The space of 4×4 matrices is clearly spanned by the basis $\mathbf{b}_j\otimes\mathbf{b}_k$. We then have the following result:

Proposition 4.3.3 *There exist scalar distributions $a_\pm$ and a_0^{mn}, $m, n = 1, 2$ so that the limit Wigner distribution matrix can be decomposed as*

$$W(t,x,k) = \sum_{j,m=1}^{2} a_0^{jm}(t,x,k)\mathbf{b}_j(x,k)\otimes\mathbf{b}_m(x,k) \\ + a_+(t,x,k)\mathbf{b}_+(x,k)\otimes\mathbf{b}_+(x,k) + a_-(t,x,k)\mathbf{b}_-(x,k)\otimes\mathbf{b}_-(x,k). \quad (4.34)$$

The main result of this proposition is that the cross terms $\mathbf{b}_j \otimes \mathbf{b}_k$ with $\omega_j \neq \omega_k$ do not contribute to the limit W. The proof of this proposition can be found in [31] and a formal derivation in [53].

The initial conditions for the amplitudes a_j are calculated using the identity

$$A_0^{-1}(x) = \sum_{j\in J}\mathbf{b}_j(x,k)\otimes\mathbf{b}_j(x,k).$$

Then (4.28) implies that $a_0^{12}(0,x,k) = a_0^{21}(0,x,k) = 0$ and

$$a_0^{jj}(0,x,k) = a_\pm(0,x,k) = |\chi(x)|^2 f(k), \quad j = 1, 2. \quad (4.35)$$

4.3.5 Mode Decomposition and Refocusing

We can use the above result to recast (4.29) as

$$\mathbf{u}^B(\xi;x_0) = (F(T,\cdot;x_0) * \mathbf{S})(\xi), \quad (4.36)$$

where

$$F(T,\xi;x_0) = \sum_{m,n=1}^{2}\int_{\mathbb{R}^3} e^{ik\cdot\xi} a_0^{mn}(T,x_0;k)\Gamma\mathbf{b}_m(x_0,k)\otimes\mathbf{b}_n(x_0,k)A_0(x_0)\Gamma\frac{dk}{(2\pi)^3} \\ + \int_{\mathbb{R}^3} e^{ik\cdot\xi} a_+(T,x_0;k)\Gamma\mathbf{b}_+(x_0,k)\otimes\mathbf{b}_+(x_0,k)A_0(x_0)\Gamma\frac{dk}{(2\pi)^3} \\ + \int_{\mathbb{R}^3} e^{ik\cdot\xi} a_-(T,x_0;k)\Gamma\mathbf{b}_-(x_0,k)\otimes\mathbf{b}_-(x_0,k)A_0(x_0)\Gamma\frac{dk}{(2\pi)^3}. \quad (4.37)$$

This expression can be used to assess the quality of the refocusing. When $F(T,\xi;x_0)$ has a narrow support in ξ, refocusing is good. When its support in ξ grows larger, its quality degrades. The spatial decay of the kernel $F(t,\xi;x_0)$ in ξ is directly related to the smoothness in k of its Fourier transform in ξ:

$$\hat{F}(T,k;x_0) = \sum_{m,n=1}^{2} a_0^{mn}(T,x_0;k)\Gamma\mathbf{b}_m(x_0,k)\otimes\mathbf{b}_n(x_0,k)A_0(x_0)\Gamma\frac{dk}{(2\pi)^3} \\ +\Gamma\left[a_+(T,x_0;k)\mathbf{b}_+(x_0,k)\otimes\mathbf{b}_+(x_0,k) + a_-(T,x_0;k)\mathbf{b}_-(x_0,k)\otimes\mathbf{b}_-(x_0,k)\right]A_0(x_0)\Gamma.$$

Namely, for F to decay in ξ, one needs $\hat{F}(k)$ to be smooth in k. However, the eigenvectors b_j are singular at $k = 0$ as can be seen from the explicit expressions (4.32). Therefore, a priori $\hat{F}$ is not smooth at $k = 0$. This means that in order to obtain good refocusing one needs the original signal to have no low frequencies: $\hat{S}(k) = 0$ near $k = 0$. Low frequencies in the initial data will not refocus well.

We can further simplify (4.36)-(4.37) is we assume that the initial condition is irrotational. Taking Fourier transform of both sides in (4.36), we obtain that

$$\hat{\mathbf{u}}^B(k;x_0) = \sum_{j,n\in J} a_j(T,x_0,k)\hat{S}_n(k)(A_0(x_0)\Gamma b_n(x_0,k)\cdot b_j(x_0,k))\Gamma b_j(x_0,k) \quad (4.38)$$

where we have defined

$$\hat{\mathbf{S}}(k) = \sum_{n\in J} \hat{S}_n(k)\mathbf{b}_n(x_0, k). \tag{4.39}$$

Irrotationality of the initial condition means that $\hat{S}_1$ and $\hat{S}_2$ identically vanish, or equivalently that

$$\mathbf{S}(x) = \begin{pmatrix} \nabla\phi(x) \\ p(x) \end{pmatrix} \tag{4.40}$$

for some pressure $p(x)$ and potential $\phi(x)$. Remarking that $\Gamma\mathbf{b}_\pm = -\mathbf{b}_\mp$ and by irrotationality that $(A_0(x_0)\hat{\mathbf{S}}(k) \cdot \mathbf{b}_{1,2}(k)) = 0$, we use (4.33) to recast (4.38) as

$$\hat{\mathbf{u}}^B(k; x_0) = a_-(T, x_0, k)\hat{S}_+(k)\mathbf{b}_+(x_0, k) + a_+(T, x_0, k)\hat{S}_-(k)\mathbf{b}_-(x_0, k). \tag{4.41}$$

Decomposing the initial condition $\mathbf{S}(x)$ as

$$\mathbf{S}(x) = \mathbf{S}_+(x) + \mathbf{S}_-(x), \qquad \text{such that} \quad \hat{\mathbf{S}}_\pm(k) = \hat{S}_\pm(k)\mathbf{b}_\pm(x_0, k),$$

the back-propagated signal takes the form

$$\mathbf{u}^B(\xi; x_0) = (\hat{a}_-(T, x_0, \cdot) * \mathbf{S}_+(\cdot))(\xi) + (\hat{a}_+(T, x_0, \cdot) * \mathbf{S}_-(\cdot))(\xi) \tag{4.42}$$

where $\hat{a}_\pm$ is the Fourier of $a_\pm$ in k. This form is much more tractable than (4.36)-(4.37). It is also almost as general. Indeed, rotational modes do not propagate in the high frequency regime. Therefore, they are exactly back-propagated when $\chi(x_0) = 1$ and $f(x) = \delta(x)$, and not back-propagated at all when $\chi(x_0) = 0$. All the refocusing properties are thus captured by the amplitudes $a_\pm(T, x_0, k)$. Their evolution equation characterizes how waves propagate in the medium and their initial conditions characterize the recording array.

4.3.6 Homogeneous Media

In homogeneous media with $c(x) = c_0$ the amplitudes $a_\pm(T, x, k)$ satisfy the free transport equation [31, 53]

$$\frac{\partial a_\pm}{\partial t} \pm c_0\hat{k} \cdot \nabla_x a_\pm = 0 \tag{4.43}$$

with initial data $a_\pm(0, x, k) = |\chi(x)|^2 f(k)$ as in (4.35). They are therefore given by

$$a_\pm(t, x_0, k) = |\chi(x_0 \mp c_0\hat{k}t)|^2 \hat{f}(k). \tag{4.44}$$

These amplitudes become more and more singular in k as time grows since their gradient in k grows linearly with time. The corresponding kernel $F = F_H$ decays therefore more slowly in ξ as time grows. This implies that the quality of the refocusing degrades with time. For sufficiently large times, all the energy has left the domain Ω (assumed to be bounded), and the coefficients $a_\pm(t, x_0, k)$ vanish. Therefore the back-propagated signal $\mathbf{u}^B(\xi; x_0)$ also vanishes, which means that there is no refocusing at all. The same conclusions could also be drawn by analyzing (4.14) directly in a homogeneous medium. This is the situation in the numerical experiment presented in Fig. 4.3: in a homogeneous medium, the back-propagated signal would vanish.

4.3.7 Heterogeneous Media and Radiative Transport Regime

The results of the preceding sections show how the back-propagated signal $\mathbf{u}^B(\xi; x_0)$ is related to the propagating modes $a_\pm(T, x_0, k)$ of the Wigner matrix $W(T, x_0, k)$. The form assumed by the modes $a_\pm(T, x_0, k)$, and in particular their smoothness in k, will depend on the hypotheses we make on the underlying medium; i.e., on the density $\rho(x)$ and compressibility $\kappa(x)$ that appear in the matrix $A(x)$. We have seen that partial measurements in homogeneous media yield poor refocusing properties. We now show that refocusing is much better in random media.

We consider here the radiative transport regime, also known as weak coupling limit. There, the fluctuations in the physical parameters are weak and vary on a scale comparable to the scale of the initial condition. Density and compressibility assume the form

$$\rho(x) = \rho_0 + \sqrt{\varepsilon}\rho_1(\frac{x}{\varepsilon}) \qquad \text{and} \qquad \kappa(x) = \kappa_0 + \sqrt{\varepsilon}\kappa_1(\frac{x}{\varepsilon}). \tag{4.45}$$

The functions ρ_1 and κ_1 are assumed to be mean-zero spatially homogeneous processes. The average (with respect to realizations of the medium) of the propagating amplitudes $a_\pm$, denoted by $\bar{a}_\pm$, satisfy in the high frequency limit $\varepsilon \to 0$ a radiative transfer equation (RTE), which is a linear Boltzmann equation of the form

$$\begin{aligned} &\frac{\partial \bar{a}_\pm}{\partial t} \pm c_0 \hat{k} \cdot \nabla_x \bar{a}_\pm = \int_{\mathbb{R}^3} \sigma(k,p)(\bar{a}_\pm(t,x,p) - \bar{a}_\pm(t,x,k))\delta(c_0(|k| - |p|))dp \\ &\bar{a}_\pm(0,x,k) = |\chi(x)|^2 \hat{f}(k). \end{aligned} \tag{4.46}$$

The scattering coefficient $\sigma(k,p)$ depends on the power spectra of ρ_1 and κ_1. We refer to [53] for the details of the derivation and explicit form of $\sigma(k,p)$. The above result remains formal for the wave equation and requires averaging over the realizations of the random medium although this is not necessary in the physical and numerical time reversal experiments. A rigorous derivation of the linear Boltzmann equation (which also requires averaging over realizations) has only been obtained for the Schrödinger equation; see [23, 56]. Nevertheless, the above result formally characterizes the filter $F(T, \xi; x_0)$ introduced in (4.37) and (4.42).

The transport equation (4.46) has a smoothing effect best seen in its integral formulation. Let us define the total scattering coefficient $\Sigma(k) = \int_{\mathbb{R}^3} \sigma(k,p)\delta(c_0(|k| - |p|))dp$. Then the transport equation (4.46) may be rewritten as

$$\begin{aligned} &\bar{a}_\pm(t,x,k) = \bar{a}_\pm(0, x \mp c_0\hat{k}t, k)e^{-\Sigma(k)t} \\ &+\frac{|k|^2}{c_0}\int_0^t ds \int_{S^2} \sigma(k, |k|\hat{\mathbf{p}})\bar{a}_\pm(s, x \mp c_0(t-s)\hat{k}, |k|\hat{\mathbf{p}})e^{-\Sigma(k)(t-s)}d\Omega(\hat{\mathbf{p}}). \end{aligned} \tag{4.47}$$

Here $\hat{\mathbf{p}} = p/|p|$ is the unit vector in direction of p and $d\Omega(\hat{\mathbf{p}})$ is the surface element on the sphere S^2. The first term in (4.47) is the ballistic part that undergoes no scattering. It has no smoothing effect, and, moreover, if $a(0,x,k)$ is not smooth in x, as may be the case for (4.35), the discontinuities in x translate into discontinuities in k at later times as in (4.44) in a homogeneous medium. However, in contrast to the homogeneous medium case, the ballistic term decays exponentially in time, and does not affect the refocused signal for sufficiently long times $t \gg 1/\Sigma$. The second term in (4.47) exhibits a smoothing effect. Namely the operator $\mathcal{L}g$ defined by

$$\mathcal{L}g(t,x,k) = \frac{|k|^2}{c_0}\int_0^t ds \int_{S^2} \sigma(k, |k|\hat{\mathbf{p}})g(s, x \mp c_0(t-s)\hat{k}, |k|\hat{\mathbf{p}})e^{-\Sigma(k)(t-s)}d\Omega(\hat{\mathbf{p}})$$

is regularizing, in the sense that the function $\tilde{g} = \mathcal{L}g$ has at least 1/2-more derivatives than g (in some Sobolev scale). The precise formulation of this smoothing property is given by the averaging lemmas [34, 51] and will not be dwelt upon here. Iterating (4.47) n times we obtain

$$\bar{a}_\pm(t,x,k) = a^0_\pm(t,x,k) + a^1_\pm(t,x,k) + \cdots + a^n_\pm(t,x,k) + \mathcal{L}^{n+1}\bar{a}_\pm(t,x,k). \tag{4.48}$$

The terms $a^0_\pm, \dots, a^n_\pm$ are given by

$$a^0_\pm(t,x,k) = \bar{a}_\pm(0, x \mp c_0\hat{k}t, k)e^{-\Sigma(k)t}, \quad a^j_\pm(t,x,k) = \mathcal{L}a^{j-1}_\pm(t,x,k).$$

They describe, respectively, the contributions from waves that do not scatter, scatter once, twice, ... It is straightforward to verify that all these terms decay exponentially in time and are negligible for times $t \gg 1/\Sigma$. The last term in (4.48) has at least $n/2$ more derivatives than the initial data a_0, or the solution (4.44) of the homogeneous transport equation. This leads to a faster decay in ξ of the Fourier transforms $\hat{a}_\pm(T, x_0, \xi)$ of $a_\pm(T, x_0, k)$ in k. This gives a qualitative explanation as to why refocusing is better in heterogeneous media than in homogeneous media. A more quantitative answer requires to solve the transport equation (4.46).

4.3.8 Diffusion Regime

It is known for times t much longer than the scattering mean free time $\tau_{sc} = 1/\Sigma$ and distances of propagation L very large compared to $l_{sc} = c_0\tau_{sc}$ that solutions to the radiative transport equation (4.46) can be approximated by solutions to a diffusion equation, provided that $c(x) = c_0$ is independent of x [19, 48]. More precisely, we let $\delta = l_{sc}/L \ll 1$ be a small parameter and rescale time and space variables as $t \to t/\delta^2$ and $x \to x/\delta$. In this limit, the wave direction is completely randomized so that

$$\bar{a}_+(t,x,k) \approx \bar{a}_-(t,x,k) \approx a(t,x,|k|),$$

where a solves

$$\begin{aligned} &\frac{\partial a(t,x,|k|)}{\partial t} - D(|k|)\Delta_x a(t,x,|k|) = 0, \\ &a(0,x,|k|) = |\chi(x)|^2 \frac{1}{4\pi|k|^2}\int_{\mathbb{R}^3} \hat{f}(q)\delta(|q| - |k|)dq. \end{aligned} \tag{4.49}$$

The diffusion coefficient $D(|k|)$ may be expressed explicitly in terms of the scattering coefficient $\sigma(k,p)$ and hence related to the power spectra of ρ_1 and κ_1. We refer to [53] for the details. For instance, let us assume for simplicity that the density is not fluctuating, $\rho_1 \equiv 0$, and that the compressibility fluctuations are delta-correlated, so that $\mathbb{E}\{\hat{\kappa}_1(p)\hat{\kappa}_1(q)\} = \kappa_0^2\hat{R}_0\delta(p+q)$. Then we have

$$\sigma(k,p) = \frac{\pi c_0^2|k|^2\hat{R}_0}{2}, \qquad \Sigma(|k|) = 2\pi^2 c_0|k|^4\hat{R}_0 \tag{4.50}$$

and

$$D(|k|) = \frac{c_0^2}{3\Sigma(|k|)} = \frac{c_0}{6\pi^2|k|^4\hat{R}_0} \tag{4.51}$$

Let us assume that there are no initial rotational modes, so that the source $\mathbf{S}(x)$ is decomposed as in (4.40). Using (4.41), we obtain that

$$\hat{\mathbf{u}}^B(k;x_0) = a(T, x_0, |k|)\hat{\mathbf{S}}(k). \tag{4.52}$$

When $f(x)$ is isotropic so that $\hat{f}(k) = \hat{f}(|k|)$, and the diffusion coefficient is given by (4.51), the solution of (4.49) takes the form

$$a(T, x_0, |k|) = \hat{f}(|k|)\Big(\frac{3\pi|k|^4\hat{R}_0}{2c_0T}\Big)^{3/2}\int_{\mathbb{R}^3} \exp\Big(-\frac{3\pi^2|k|^4\hat{R}_0|x_0-y|^2}{2c_0T}\Big)|\chi(y)|^2 dy. \qquad (4.53)$$

When $f(x) = \delta(x)$, and $\Omega = \mathbb{R}^3$, so that $\chi(x) \equiv 1$, we retrieve $a(T, x_0, k) \equiv 1$, hence the refocusing is perfect. When only partial measurement is available, the above formula indicates how the frequencies of the initial pulse are filtered by the one-step time reversal process. Notice that both the low and high frequencies are damped. The reason is that low frequencies scatter little from the underlying medium so that it takes a long time for them to be randomized. High frequencies strongly scatter with the underlying medium and consequently propagate little so that the signal that reaches the recording array Ω is small unless recorders are also located at the source point: $x_0 \in \Omega$. In the latter case they are very well measured and back-propagated although this situation is not the most interesting physically. Expression (4.53) may be generalized to other power spectra of medium fluctuations in a straightforward manner using the formula for the diffusion coefficient in [53].

4.3.9 Numerical Results

The numerical results in Fig. 4.3 show that some signal refocuses at the location of the initial source after the time reversal procedure. Based on the above theory however, we do not expect the refocused signal to have exactly the same shape as the original one. Since the location of the initial source belongs to the recording array ($\chi(x_0) = 1$) in our simulations, we expect from our theory that high frequencies will refocus well but that low frequencies will not. This

Figure 4.4: Zoom of the initial source and the refocused signal for the numerical experiment of Fig. 4.3.

is confirmed by the numerical results in Fig. 4.4, where a zoom in the vicinity of $x_0 = \mathbf{0}$ of the initial source and refocused signal are represented. Notice that the numerical simulations are presented here only to help in the understanding of the refocusing theory and do not aim at reproducing the theory in a quantitative manner. The random fluctuations are quite strong in our numerical simulations and it is unlikely that the diffusive regime will be valid. The refocused signal on the right figure looks however like a high-pass filter of the signal in the left figure, as expected from theory.

4.4 Refocusing of Classical Waves

The theory presented in section 4.3 provides a quantitative explanation for the results observed in time reversal physical and numerical experiments. However, the time reversal procedure is by no means necessary to obtain refocusing. Time reversal is associated with the specific choice (4.12) for the matrix Γ in the preceding section, which reverses the direction of the acoustic velocity and keeps pressure unchanged. Other choices for Γ are however possible. When nothing is done at time T, i.e., when we choose $\Gamma = I$, no refocusing occurs as one might expect. It turns out that $\Gamma = I$ is more or less the only choice of a matrix that prevents some sort of refocusing. Section 4.4.1 presents the theory of refocusing for acoustic waves, which is corroborated by numerical results presented in Section 4.4.2. Sections 4.4.3 and 4.4.4 generalize the theory to other linear hyperbolic systems.

4.4.1 General Refocusing of Acoustic Waves

In one-step time reversal, the action of the "intelligent" array is captured by the choice of the signal processing matrix Γ in (4.13). Time reversal is characterized by Γ given in (4.12). A passive array is characterized by $\Gamma = I$. This section analyzes the role of other choices for Γ, which we let depend on the receiver location so that each receiver may perform its own kind of signal processing.

The signal after time reversal is still given by (4.13), where $\Gamma(y')$ is now arbitrary. At time $2T$, after back-propagation, we are free to multiply the signal by an arbitrary invertible matrix to analyze the signal. It is convenient to multiply the back-propagated signal by the matrix $\Gamma_0 = \text{Diag}(-1, -1, -1, 1)$ as in classical time reversal. The reconstruction formula (4.15) in the localized source limit is then replaced by

$$\mathbf{u}^B(\xi; x_0) = \int_{\mathbb{R}^9} \Gamma_0 G(T, x_0 + \varepsilon\xi; y)\Gamma(y')G(T, y'; x_0 + \varepsilon z)\chi(y, y')\mathbf{S}(z)dydy'dz \tag{4.54}$$

with $\chi(y, y')$ defined by (4.16). To generalize the results of section 4.3, we need to define an appropriate adjoint Green's matrix G_*. As before, this will allow us to remove the matrix Γ between the two Green's matrices in (4.54) and to interchange the order of points in the second Green's matrix. We define the new adjoint Green's function $G_*(t, x; y)$ as the solution to

$$\begin{aligned} &\frac{\partial G_*(t, x; y)}{\partial t} A(x) + \frac{\partial G_*(t, x; y)}{\partial x^j} D^j = 0 \\ &G_*(0, x; y) = \Gamma(x)\Gamma_0 A^{-1}(x)\delta(x - y). \end{aligned} \tag{4.55}$$

Following the steps of section 4.3.3, we show that

$$G_*(t, x, y) = \Gamma(y)G(t, y; x)A^{-1}(x)\Gamma_0. \tag{4.56}$$

The only modification compared to the corresponding derivation of (4.18) is to multiply (4.19) on the left by $\Gamma(x)$ and on the right by Γ_0 so that $\Gamma(y)$ appears on the left in (4.20). The re-transmitted signal may now be recast as

$$\mathbf{u}^B(\xi; x_0) = \int_{\mathbb{R}^9} \Gamma_0 G(T, x_0 + \varepsilon\xi; y)G_*(T, x_0 + \varepsilon z; y')\Gamma_0^{-1} A(x_0 + \varepsilon z)\chi(y, y')\mathbf{S}(z)dydy'dz. \tag{4.57}$$

Therefore the only modification in the expression for the re-transmitted signal compared to the time reversed signal (4.22) is in the initial data for (4.55), which is the only place where the matrix $\Gamma(x)$ appears.

The analysis in Sections 4.3.3-4.3.7 requires only minor changes, which we now outline. The back-propagated signal may still be expressed in term of the Wigner distribution (compare to (4.27))

$$\mathbf{u}^B(\xi; x_0) = \int_{\mathbb{R}^6} e^{ik\cdot(\xi - z)}\Gamma_0 W_\varepsilon(T, x_0 + \varepsilon\frac{z + \xi}{2}, k)\Gamma_0 A(x_0 + \varepsilon z)\mathbf{S}(z)\frac{dzdk}{(2\pi)^3}. \tag{4.58}$$

The Wigner distribution is defined as before by (4.25) and (4.26). The function Q is defined as before as the solution of (4.4) with initial data $Q(0, x; q) = \chi(x)e^{iq\cdot x/\varepsilon}I$, while Q_* solves (4.17) with the initial data $Q_*(0, x; q) = \Gamma(x)\Gamma_0 A^{-1}(x)\chi(x)e^{-iq\cdot x/\varepsilon}$. The initial Wigner distribution is now given by

$$W(0, x, k) = |\chi(x)|^2\Gamma(x)\Gamma_0 A^{-1}(x)\hat{f}(k). \tag{4.59}$$

Lemma 4.3.1 and Proposition 4.3.2 also hold, and we obtain the analog of (4.29)

$$\mathbf{u}(\xi; x_0) = \int_{\mathbb{R}^6} e^{ik\cdot(\xi - z)} \Gamma_0 W(T, x_0, k) \Gamma_0 A_0(x_0) \mathbf{S}(z) dz dk. \tag{4.60}$$

The limit Wigner distribution $W(T, x_0, k)$ admits the mode decomposition (4.34) as before. If we assume that the source $\mathbf{S}(x)$ has the form (4.40) so that no rotational modes are present initially, we recover the refocusing formula (4.41):

$$\hat{\mathbf{u}}^B(k; x_0) = a_-(T, x_0, k)\hat{S}_+(k) b_+(x_0, k) + a_+(T, x_0, k)\hat{S}_-(k) b_-(x_0, k). \tag{4.61}$$

The initial conditions for the amplitudes $a_\pm$ are replaced by

$$\begin{aligned} a_\pm(0, x, k) &= \mathrm{Tr}\left[A_0(x) W(0, x, k) A_0(x) b_\pm(x_0, k) b_\pm^*(x_0, k)\right] \\ &= |\chi(x)|^2 \hat{f}(k)(A_0(x)\Gamma(x) b_\mp(x, k) \cdot b_\pm(x, k)). \end{aligned} \tag{4.62}$$

Observe that when $\Gamma(x) = \Gamma_0$, we get back the results of Section 4.3.7. When the signal is not changed at the array, so that $\Gamma = I$, the coefficients $a_\pm(0, x, k) \equiv 0$ by orthogonality (4.33) of the eigenvectors $\mathbf{b}_j$. We thus obtain that no refocusing occurs when the "intelligent" array is replaced by a passive array, as expected physically.

Another interesting example is when only pressure p is measured, so that the matrix $\Gamma = \mathrm{Diag}(0, 0, 0, 1)$. Then the initial data is

$$a_\pm(0, x, k) = \frac{1}{2}|\chi(x)|^2 \hat{f}(k),$$

which differs by a factor 1/2 from the full time reversal case (4.35). Therefore the re-transmitted signal $\mathbf{u}^B$ also differs only by a factor 1/2 from the latter case, and the quality of refocusing as well as the shape of the re-propagated signal are exactly the same. The same observation applies to the measurement and reversal of the acoustic velocity only, which corresponds to the matrix $\Gamma = \mathrm{Diag}(-1, -1, -1, 0)$. The factor 1/2 comes from the fact that only the potential energy or the kinetic energy is measured in the first and second cases, respectively. For high frequency acoustic waves, the potential and kinetic energies are equal, hence the factor 1/2. We can also verify that when only the first component of the velocity field is measured so that $\Gamma = \mathrm{Diag}(-1, 0, 0, 0)$, the initial data is

$$a_\pm(0, x, k) = |\chi(x)|^2 \hat{f}(k) \frac{k_1^2}{2|k|^2}. \tag{4.63}$$

As in the time reversal setting of Section 4.3, the quality of the refocusing is related to the smoothness of the amplitudes $a_\pm$ in k. In a homogeneous medium they satisfy the free transport equation (4.43), and are given by

$$a_\pm(t, x, k) = |\chi(x - c_0\hat{k}t)|^2 \hat{f}(k)(A_0(x - c_0\hat{k}t)\Gamma(x - c_0\hat{k}t) b_\mp(x - c_0\hat{k}t, k) \cdot b_\pm(x - c_0\hat{k}t, k)).$$

Once again, we observe that in a uniform medium $a_\pm$ become less regular in k as time grows, thus refocusing is poor.

The considerations of Section 4.3.7 show that in the radiative transport regime the amplitudes $a_\pm$ become smoother in k also with initial data given by (4.62). This leads to a better refocusing as explained in Section 4.3.5. Let us assume that the diffusion regime of Section 4.3.8 is valid and that the kernel f is isotropic $\hat{f}(k) = \hat{f}(|k|)$. This requires in particular that $A_0(x)$ be independent of x. We obtain that $a_\pm(T, x_0, k) = \tilde{a}(T, x_0, |k|)$, thus the refocusing formula (4.61) reduces to

$$\hat{\mathbf{u}}^B(k; x_0) = \tilde{a}(T, x_0, |k|)\hat{\mathbf{S}}(k). \tag{4.64}$$

The difference with the case treated in Section 4.3.8 is that $\tilde{a}(T, x, |k|)$ solves the diffusion equation (4.49) with new initial conditions given by

$$\begin{aligned}\tilde{a}(0, x, |k|) &= \frac{|\chi(x)|^2}{4\pi|k|^2} \int_{\mathbb{R}^3} \hat{f}(|q|)(A_0\Gamma(x)\mathbf{b}_-(q) \cdot \mathbf{b}_+(q))\delta(|q| - |k|)dq \\ &= \frac{|\chi(x)|^2}{4\pi|k|^2} \int_{\mathbb{R}^3} \hat{f}(|q|)(A_0\Gamma(x)\mathbf{b}_+(q) \cdot \mathbf{b}_-(q))\delta(|q| - |k|)dq.\end{aligned} \tag{4.65}$$

When only the first component of the velocity field is measured, as in (4.63), the initial data for $\tilde{a}$ is

$$\tilde{a}(0, x, |k|) = \frac{1}{6}|\chi(x)|^2 \hat{f}(|k|).$$

Therefore even time reversing only one component of the acoustic velocity field produces a re-propagated signal that is equal to the full re-propagated field up to a constant factor.

More generally, we deduce from (4.65) that a detector at x will contribute some refocusing for waves with wavenumber $|k|$ provided that

$$\int_{S^2} \hat{f}(|k|\hat{q})(A_0\Gamma(x)\mathbf{b}_\mp(\hat{q}) \cdot \mathbf{b}_\pm(\hat{q}))d\Omega(\hat{q}) \neq 0.$$

When $f(x) = f(|x|)$ is radial, this property becomes independent of the wavenumber $|k|$ and reduces to $\int_{S^2}(A_0\Gamma(x)\mathbf{b}_\mp(\hat{q}) \cdot \mathbf{b}_\pm(\hat{q}))d\Omega(\hat{q}) \neq 0$.

4.4.2 Numerical Results

Let us come back to the numerical results presented in Fig. 4.3 and 4.4. We now consider two different processings at the recording array. The first array is passive, corresponding to $\Gamma = I$, and the second array only measures pressure so that $\Gamma = \text{Diag}(0, 0, 0, 1)$. The zoom in the vicinity of $x_0 = \mathbf{0}$ of the "refocused" signals is given in Fig. 4.5. The left figure shows

Figure 4.5: Zoom of the refocused signals for the numerical experiment of Fig. 4.3 with processing $\Gamma = I$ (left), with a maximal amplitude of roughly $4\,10^{-3}$ and $\Gamma = \text{Diag}(0, 0, 0, 1)$ (right), with a maximal amplitude of roughly 0.035.

no refocusing, in accordance with physical intuition and theory. The right figure shows that refocusing indeed occurs when only pressure in recorded (and its time derivative is set to 0 in the solution of the wave equation presented in the appendix). Notice also that the refocused signal is roughly one half the one obtained in Fig. 4.4 as predicted by theory.

4.4.3 Refocusing of Other Classical Waves

The preceding sections deal with the refocusing of acoustic waves. The theory can however be extended to more complicated linear hyperbolic systems of the form (4.4) with $A(x)$ a positive definite matrix, D^j symmetric matrices, and $\mathbf{u} \in \mathbb{C}^m$. These include electromagnetic and elastic waves. Their explicit representation in the form (4.4) and expressions for the matrices $A(x)$ and D^j in these cases may be found in [53]. For instance, the Maxwell equations

$$\begin{aligned}\frac{\partial E}{\partial t} &= \frac{1}{\epsilon(x)} \text{curl } H \\ \frac{\partial H}{\partial t} &= -\frac{1}{\mu(x)}\text{curl } E\end{aligned}$$

may be written in the form (4.4) with $\mathbf{u} = (E, H) \in \mathbb{C}^6$ and the matrix

$$A(x) = \mathrm{Diag}(\epsilon(x), \epsilon(x), \epsilon(x), \mu(x), \mu(x), \mu(x)).$$

Here $\epsilon(x)$ is the dielectric constant (not to be confused with the small parameter ε), and $\mu(x)$ is the magnetic permeability. The 6×6 dispersion matrix $L(x, k)$ for the Maxwell equations is given by

$$L(x,k) = -\begin{pmatrix} 0 & 0 & 0 & 0 & -k_3/\epsilon(x) & k_2/\epsilon(x) \\ 0 & 0 & 0 & k_3/\epsilon(x) & 0 & -k_1/\epsilon(x) \\ 0 & 0 & 0 & -k_2/\epsilon(x) & k_1/\epsilon(x) & 0 \\ 0 & k_3/\mu(x) & -k_2/\mu(x) & 0 & 0 & 0 \\ -k_3/\mu(x) & 0 & k_1/\mu(x) & 0 & 0 & 0 \\ k_2/\mu(x) & -k_1/\mu(x) & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Generalization of our results for acoustic waves to such general systems is quite straightforward so we concentrate only on the modifications that need be made. The time reversal procedure is exactly the same as before: a signal propagates from a localized source, is recorded, processed as in (4.13) with a general matrix $\Gamma(y')$, and re-emitted into the medium. The retransmitted signal is given by (4.54). Furthermore, the equation for the adjoint Green's matrix (4.55), the definition of the Wigner transform in Section 4.3.4, and the expression (4.60) for the re-propagated signal still hold.

The analysis of the re-propagated signal is reduced to the study of the Wigner distribution, which is now modified. The mode decomposition must be generalized. We recall that

$$L(x,k) = A_0^{-1}(x) k_j D^j$$

is the $m \times m$ dispersion matrix associated with the hyperbolic system (4.4). Since $L(x, k)$ is symmetric with respect to the inner product $\langle \mathbf{u}, \mathbf{v} \rangle_{A_0} = (A_0 \mathbf{u} \cdot \mathbf{v})$, its eigenvalues are real and its eigenvectors form a basis. We assume the existence of a time reversal matrix Γ_0 such that (4.21) holds with $\Gamma = \Gamma_0$ and such that $\Gamma_0^2 = I$. For example, for electromagnetic waves $\Gamma_0 = \mathrm{Diag}(1, 1, 1, -1, -1, -1)$. Then the spectrum of L is symmetric about zero and the eigenvalues $\pm\omega^\alpha$ have the same multiplicity. We assume in addition that L is isotropic so that its eigenvalues have the form $\omega_\pm^\alpha(x, k) = \pm c^\alpha(x)|k|$, where $c_\alpha(x)$ is the speed of mode α. We denote by r_α their respective multiplicities, assumed to be independent of x and k for $k \neq 0$. The matrix L has a basis of eigenvectors $\mathbf{b}_\pm^{\alpha,j}(x, k)$ such that

$$L(x,k)\mathbf{b}_\pm^{\alpha,j}(x,k) = \pm\omega^\alpha(x,k)\mathbf{b}_\pm^{\alpha,j}(x,k), \quad j = 1, \ldots, r_\alpha,$$

and $b_\pm^{\alpha,j}$ form an orthonormal set with respect to the inner product $\langle , \rangle_{A_0}$. The different ω_α correspond to different types of waves (modes). Various indices $1 \le j \le r_\alpha$ refer to different polarizations of a given mode. The eigenvectors $b_+^{\alpha,j}$ and $b_-^{\alpha,j}$ are related by

$$\Gamma_0 \mathbf{b}_+^{\alpha,j}(x,k) = \mathbf{b}_-^{\alpha,j}(x,k), \quad \Gamma_0 \mathbf{b}_-^{\alpha,j}(x,k) = \mathbf{b}_+^{\alpha,j}(x,k). \tag{4.66}$$

Proposition 4.3.3 is then generalized as follows [31, 53]:

Proposition 4.4.1 *There exist scalar functions $a_\pm^{\alpha,jm}(t, x, k)$ such that*

$$W(t,x,k) = \sum_{\pm,\alpha,j,m} a_\pm^{\alpha,jm}(t,x,k)\mathbf{b}_\pm^{\alpha,j}(x,k) \otimes \mathbf{b}_\pm^{\alpha,m}(x,k). \tag{4.67}$$

Here the sum runs over all possible values of $\pm$, α, and $1 \le j, m \le r_\alpha$.

The main content of this proposition is again that the cross terms $\mathbf{b}_\pm^{\alpha,j}(x,k) \otimes \mathbf{b}_\mp^{\beta,m}(x,k)$ do not contribute, as well as the terms $\mathbf{b}_\pm^{\alpha,j}(x,k) \otimes \mathbf{b}_\pm^{\alpha',m}(x,k)$ when $\alpha \neq \alpha'$. This is because modes propagating with different speeds do not interfere constructively in the high frequency limit.

We may now insert expression (4.67) into (4.60) and obtain the following generalization of (4.61)

$$\hat{\mathbf{u}}^B(k;x_0) = \sum_{\alpha,j,m} \Big[a_-^{\alpha,mj}(T,x_0,k) \hat{S}_+^{\alpha,j}(x_0,k) \mathbf{b}_+^{\alpha,m}(x_0,k) + a_+^{\alpha,mj}(T,x_0,k) \hat{S}_-^{\alpha,j}(x_0,k) \mathbf{b}_-^{\alpha,m}(x_0,k) \Big], \tag{4.68}$$

where $\hat{S}_\pm^{\alpha,j}(k) = (A(x_0)\hat{\mathbf{S}}(k) \cdot \mathbf{b}_\pm^{\alpha,j}(x_0,k))$. This formula tells us that only the modes that are present in the initial source ($\hat{S}_\pm^{\alpha,j}(k) \neq 0$) will be present in the back-propagated signal but possibly with a different polarization, that is, $j \neq m$.

The initial conditions for the modes $a_\pm^{\alpha,jm}$ are given by

$$a_\pm^{\alpha,jm}(0,x,k) = |\chi(x)|^2 \hat{f}(k) (A(x)\Gamma(x) \mathbf{b}_\mp^{\alpha,m}(x,k) \cdot \mathbf{b}_\pm^{\alpha,j}(x,k)), \tag{4.69}$$

which generalizes (4.62). When $\Gamma(x) \equiv I$, we again obtain that $a_\pm^{\alpha,jm}(0,x,k) \equiv 0$, i.e., there is no refocusing as physically expected. When $\Gamma(x) \equiv \Gamma_0$, we have for all α that

$$a_\pm^{\alpha,jm}(0,x,k) = |\chi(x)|^2 \hat{f}(k) \delta_{jm}.$$

In a uniform medium the amplitudes $a_\pm^{\alpha,jm}$ satisfy an uncoupled system of free transport equations (4.43):

$$\frac{\partial a_\pm^{\alpha,jm}}{\partial t} \pm c_\alpha \hat{k} \cdot \nabla_x a_\pm^{\alpha,jm} = 0, \tag{4.70}$$

which have no smoothing effect, and hence refocusing in a homogeneous medium is still poor. When $f(x) = \delta(x)$ and $\Omega = \mathbb{R}^3$, so that $\chi(x) \equiv 1$, we still have that $a_\pm^{\alpha,jm}(T,x_0,k) = \delta_{jm}$ and refocusing is again perfect, that is, $\mathbf{u}^B(\xi;x_0) = \mathbf{S}(\xi)$, as may be seen from (4.68).

4.4.4 The diffusive regime

The radiative transport regime holds when the matrices $A(x)$ have the form

$$A(x) = A_0(x) + \sqrt{\varepsilon} A_1\Big(\frac{x}{\varepsilon}\Big),$$

as in (4.45). Then the $r_\alpha \times r_\alpha$ coherence matrices $w_\pm^\alpha$ with entries $w_{\pm,jm}^\alpha = a_\pm^{\alpha,jm}$ satisfy a system of matrix-valued radiative transport equations (see [53] for the details) similar to (4.46). The matrix transport equations simplify considerably in the diffusive regime, such as the one considered in Section 4.3.8 when waves propagate over large distances and long times. We assume for simplicity that $A_0 = A_0(x)$ and $\Gamma = \Gamma(x)$ are independent of x. Polarization is lost in this regime, that is, $a^{\alpha,jm}(t,x,k) = 0$ for $j \neq m$ and wave energy is equidistributed over all directions. This implies that

$$a_+^{\alpha,jj}(t,x,k) = a_-^{\alpha,jj}(t,x,k) = a_\alpha(t,x,|k|)$$

so that $a^{\alpha,jj}$ is independent of $j = 1, \ldots, r_\alpha$ and of the direction $\hat{k} = k/|k|$. Furthermore, because of multiple scattering, a universal equipartition regime takes place so that

$$a_\alpha(t,x_0,|k|) = \phi(t,x_0,c_\alpha|k|), \tag{4.71}$$

where $\phi(t, x, \omega)$ solves a diffusion equation in x like (4.49) (see [53]). The diffusion coefficient $D(\omega)$ may be expressed explicitly in terms of the power spectra of the medium fluctuations [53]. Using (4.69) and (4.71), we obtain when f is isotropic the following initial data for the function ϕ

$$\phi(0, x, \omega) = \frac{1}{4\pi}|\chi(x)|^2 \int_{S^2} \frac{2}{|\alpha|} \sum_{j, \omega_\alpha > 0} \hat{f}\Big(\frac{\omega}{c_\alpha}\Big)(A_0 \Gamma \mathbf{b}_-^{\alpha,j}(\hat{k}), \mathbf{b}_+^{\alpha,j}(\hat{k})) d\Omega(\hat{k}), \tag{4.72}$$

where $|\alpha|$ is the number of non-vanishing eigenvalues of $L(x, k)$, and $d\Omega(\hat{k})$ is the Lebesgue measure on the unit sphere S^2.

Let us assume that non-propagating modes are absent in the initial source $\mathbf{S}(x)$, that is, $\hat{S}_0^j(k) = 0$ with the subscript zero referring to modes corresponding to $\omega_0 = 0$. Then (4.68) becomes

$$\hat{\mathbf{u}}(k; x_0) = \sum_{\alpha, j} \phi(T, x_0, c_\alpha |k|) \Big[\hat{S}_+^{\alpha,j}(k) \mathbf{b}_+^{\alpha,j}(x_0, k) + \hat{S}_-^{\alpha,j}(k) \mathbf{b}_-^{\alpha,j}(x_0, k)\Big]. \tag{4.73}$$

This is an explicit expression for the re-propagated signal in the diffusive regime, where ϕ solves the diffusion equation (4.49) with initial conditions (4.72).

4.5 Conclusions

This chapter presents a theory that quantitatively describes the refocusing phenomena in time reversal acoustics as well as for more general processings of acoustic and other classical waves. We show that the back-propagated signal may be expressed as the convolution (4.1) of the original source $\mathbf{S}$ with a filter F. The quality of the refocusing is therefore determined by the spatial decay of the kernel F. For acoustic waves, the explicit expression (4.37) relates F to the Wigner distribution of certain solutions of the wave equation. The decay of F is related to the smoothness in the phase space of the amplitudes $a_j(t, x, k)$ defined in Proposition 4.3.3. The latter satisfy free transport equations in homogeneous media, which sharpens the gradients of a_j and leads to poor refocusing. In contrast, the amplitudes a_j satisfy the radiative transport equation (4.46) in heterogeneous media, which has a smoothing effect. This leads to a rapid spatial decay of the filter F and a better refocusing. For longer times, a_j satisfies a diffusion equation. This allows for an explicit expression (4.52)-(4.53) of the time reversed signal. The same theory holds for more general waves and more general processing procedures at the recording array, which allows us to describe the refocusing of electromagnetic waves when only one component of the electric field is measured, for instance.

Bibliography

[1] F. Bailly, J.F. Clouet and J.P. Fouque, Parabolic and white noise approximation for waves in random media, SIAM Jour. Appl. Math. **56**, 1996, 1445–1470.

[2] G.Bal, G. Papanicolaou and L. Ryzhik, Radiative transport limit for the random Schrödinger equation, Nonlinearity, **15**, 2002, 513–529.

[3] G. Bal, On the self-averaging of wave energy in random media, SIAM Multiscale Model. Simul. , **2**, 2004, 398–420.

[4] G. Bal and L. Ryzhik, Time reversal for classical waves in random media, Comptes rendus de l'Académie des sciences - Série I - Mathématique, **333**, 2001, 1041–1046.

[5] G.Bal and L. Ryzhik, Time reversal and refocusing in random media, SIAM Jour. Appl. Math., **63**, 2003, 1475–1498.

[6] G. Bal, T. Komorowski and L. Ryzhik, Self-averaging of the Wigner transform in random media, Comm. Math. Phys., 2003, **242**, 81–135.

[7] G. Bal, Kinetics of scalar wave fields in random media, to appear in Wave Motion, 2005.

[8] C. Bardos and M. Fink.

[9] C. Bardos and M. Fink, Mathematical foundations of the time reversal mirror, Asymptotic Analysis, **29**, 2002, 157–182.

[10] J. Berryman, L. Borcea, G. Papanicolaou, and C. Tsogka. Imaging and time reversal in random media, J. Acoust. Soc. Am. , **112**, 2002, 1509–1522.

[11] P. Billingsley, *Convergence of Probability Measures*, Wiley, 1999.

[12] G. Blankenship and G. Papanicolaou, Stability and control of stochastic systems with wide-band noise disturbances, SIAM Jour. Appl. Math., **34**, 1978, 437-476.

[13] P. Blomgren, G. Papanicolaou, and H. Zhao. Super-Resolution in Time-Reversal Acoustics. *to appear in J. Acoust. Soc. Am.*, 2001.

[14] L. Borcea, G. Papanicolaou, C. Tsogka and J. Berryman, Imaging and time reversal in random media, Inverse Problems, **18**, 2002, 1247–1279.

[15] L. Borcea, G. Papanicolaou and C. Tsogka, Theory and applications of time reversal and interferometric imaging, Inverse Problems, **19**, 2003, S139–S164.

[16] L. Borcea, G. Papanicolaou and C. Tsogka, Interferometric array imaging in clutter, Inverse Problems, **21**, 2005, 1419–1460.

[17] L. Borcea, G. Papanicolaou and C. Tsogka, Coherent interferometry in finely layered random media, SIAM J. on Multiscale Model. Simul, in press, 2006.

[18] J. F. Clouet and J. P. Fouque. A time-reversal method for an acoustical pulse propagating in randomly layered media. *Wave Motion*, 25:361–368, 1997.

[19] R. Dautray and J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology. Vol.6*. Springer Verlag, Berlin, 1993.

[20] A. Derode, P. Roux and M. Fink, Robust Acoustic Time-Reversal With High-Order Multiple-Scattering, Phys. Rev. Lett., **75**, 1995, 4206–4209.

[21] D. R. Dowling, and D.R. Jackson, Narrow-Band performance of phase-conjugate arrays in dynamic random media, J. Acoust. Soc. Am., **91**, 1992, 3257–3277.

[22] D. Dürr, S. Goldstein and J. Lebowitz, Asymptotic motion of a classical particle in a random potential in two dimensions: Landau model, Comm. Math. Phys., **113**, 1987, 209–230.

[23] L. Erdös and H. T. Yau. Linear Boltzmann equation as the weak coupling limit of a random Schrödinger Equation. *Comm. Pure Appl. Math.*, 53(6):667–735, 2000.

[24] L. Erdös, M. Salmhofer and H.T. Yau, Quantum diffusion of the random Schrodinger evolution in the scaling limit I. The non-recollision diagrams Preprint, 2005.

[25] L. Erdös, M. Salmhofer and H.T. Yau, Quantum diffusion of the random Schrodinger evolution in the scaling limit II. The recollision diagrams, Preprint, 2005.

[26] M. Fink. Time reversed acoustics. *Physics Today*, 50(3):34–40, 1997.

[27] M. Fink. Chaos and time-reversed acoustics. *Physica Scripta*, 90:268–277, 2001.

[28] M. Fink and C. Prada. Acoustic time-reversal mirrors. Inverse Problems, **17**, 2001, R1–R38.

[29] J.-P. Fouque, La convergence en loi pour les processus à valeurs dans un espace nucléaire, Ann. Inst. H. Poincaré Probab. Statist. **20**, 1984, 225–245.

[30] J. Garnier, J.P. Fouque, G. Papanicolaou and K. Solna, *Wave Propagation and Time Reversal in Randomly Layered Media*, to appear.

[31] P. Gérard, P. A. Markowich, N. J. Mauser, and F. Poupaud, Homogenization limits and Wigner transforms, Comm. Pure Appl. Math., **50**, 1997, 323–380.

[32] I.I. Gikhman and A.V. Skorochod, *Theory of stochastic processes, vol. 3*, Springer Verlag, 1974, Berlin.

[33] D. Gilbarg and N.S. Trudinger, *Elliptic partial differential equations of second order*, Springer Verlag, 1998, Berlin.

[34] F. Golse, P.-L. Lions, B. Perthame, and R. Sentis. Regularity of the moments of the solution of a transport equation. *Journal of Functional Analysis 76*, pages 110–125, 1988.

[35] W. Hodgkiss, H. Song, W. Kuperman, T. Akal, C. Ferla, and D. Jackson. A long-range and variable focus phase-conjugation experiment in a shallow water. *J. Acoust. Soc. Am.*, 105:1597–1604, 1999.

[36] A. Ishimaru. *Wave Propagation and Scattering in Random Media*. New York: Academics, 1978.

[37] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, New York, 1991.

[38] J.B. Keller, Stochastic equations and wave propagation in random media, Proc. Sympos. Appl. Math. AMS, Vol. XVI, 1964, 145–170.

[39] H. Kesten and G. Papanicolaou, A limit theorem for turbulent diffusion, Comm. Math. Phys., **65**, 1979, 97-128.

[40] H. Kesten, G. C. Papanicolaou, A Limit Theorem for Stochastic Acceleration, Comm. Math. Phys., **78**, 1980, 19–63.

[41] R. Khasminskii, A limit theorem for solutions of differential equations with a random right hand side, Theory Probab. Appl. **11**, 390-406.

[42] S. R. Khosla and D. R. Dowling. Time-reversing array retrofocusing in noisy environments. *J. Acous. Soc. Am.*, 109(2):538–546, 2001.

[43] T. Komorowski, Diffusion approximation for the advection of particles in a a strongly turbulent random environment, Ann. Probab., **24**, 1996, 346–376.

[44] T. Komorowski and L. Ryzhik, Diffusion in a weakly random Hamiltonian flow, Comm. Math. Phys., **262**, 2006, 277–323.

[45] T. Komorowski and L. Ryzhik, The stochastic acceleration problem in two dimensions, to appear in Israel Journal of Mathematics, 2006.

[46] W. Kuperman, W. Hodgkiss, H. Song, T. Akal, C. Ferla, and D. Jackson. Phase-conjugation in the ocean. *J. Acoust. Soc. Am.*, 102:1–16, 1997.

[47] S. Kusuoka and D. Stroock, Appliactions of the Malliavin calculus, Part II, J. Fac. Sci. Univ. Tokyo, Sect. IA, Math, **32**, 1985, 1–76.

[48] E. W. Larsen and J. B. Keller. Asymptotic solution of neutron transport problems for small mean free paths. *J. Math. Phys.*, 15:75–81, 1974.

[49] P.-L. Lions and T. Paul. Sur les mesures de Wigner. *Rev. Mat. Iberoamericana*, 9:553–618, 1993.

[50] I. Mitoma, On the sample continuity of $\mathcal{S}'$-processes. J. Math. Soc. Japan, **35**, 1983, 629–636.

[51] M. Mokhtar-Kharroubi. *Mathematical Topics in Neutron Transport Theory.* World Scientific, Singapore, 1997.

[52] G. Papanicolaou and W. Kohler, Asymptotic theory of mixing stochastic ordinary differential equations, Comm. Pure Appl. Math., **27**, 1974, 641-668.

[53] L. Ryzhik, G. Papanicolaou, and J. B. Keller, Transport equations for elastic and other waves in random media, Wave Motion, **24**, 1996, 327–370.

[54] G. Papanicolaou, L. Ryzhik, and K. Solna, Statistical stability in time reversal, SIAM Jour. Appl. Math., **64**, 2004, 1133–1155.

[55] P. Sheng. *Introduction to Wave Scattering, Localization and Mesoscopic Phenomena.* Academic Press, New York, 1995.

[56] H. Spohn. Derivation of the transport equation for electrons moving through random impurities. *Jour. Stat. Phys.*, 17:385–412, 1977.

[57] D. Strook, *An Introduction to the Analysis of Paths on a Riemannian Manifold*, Math. Surv. and Monographs v. **74**, 2000.

[58] D. Strook, S. R. S. Varadhan *Multidimensional Diffusion Processes*, Berlin, Heidelberg, New York: Springer-Verlag, 1979.

[59] F. Tappert, The parabolic approximation method, Lecture notes in physics, vol. 70, *Wave propagation and underwater acoustics*, Springer-Verlag, 1977

[60] M. Taylor, *Partial differential equations, vol. 2*, New York, Springer-Verlag, 1996.

[61] B.S. White, The stochastic caustic, SIAM J. Appl. Math. **44**, 1984, 127–149.