# Marginal density expansions for diffusions and stochastic volatility, part II: Applications 

J.D. Deuschel, P.K. Friz, A. Jacquier, S. Violante<br>TU Berlin, TU and WIAS Berlin, TU Berlin, Imperial College


#### Abstract

In [17] we discussed density expansions for multidimensional diffusions $\left(X^{1}, \ldots, X^{d}\right)$, at fixed time $T$ and projected to their first $l$ coordinates, in the small noise regime. Global conditions were found which replace the well-known "not-in-cutlocus" condition known from heat-kernel asymptotics. In the present paper we discuss financial applications; these include tail and implied volatility asymptotics in some correlated stochastic volatility models. In particular, we solve a problem left open by A. Gulisashvili and E.M. Stein (2009).


Keywords: Density expansions in small noise and small time, sub-Riemannian geometry with drift, focal points, stochastic volatility, implied volatility, large strike and small time asymptotics for implied volatility

## 1 Introduction

Given a multi-dimensional diffusion process $\mathrm{X}_{t}=\left(X_{t}^{1}, \ldots, X_{t}^{d}: t \geq 0\right)$, started at $\mathrm{X}_{0}=\mathrm{x}_{0}$, we studied in [17] the behaviour of the probability density function $f=f(\mathrm{y}, t)$ of the projected (in general non-Markovian) process

$$
\mathrm{Y}_{t}:=\Pi_{l} \circ \mathrm{X}_{t}:=\left(X_{t}^{1}, \ldots, X_{t}^{l}\right)
$$

with $l \in\{1, \ldots, d\}$ fixed. This situation is typical in analysis of stochastic volatility models; Y may represent one (or several!) assets, the full process X contains additional stochastic volatility components (and also stochastic rates, if desired). Basket models, in the spirit of [2, 3] can also be fitted in this framework. Both short time asymptotics and tail asymptotics, in presence of suitable scaling properties of the model, can be derived from the small noise problem

$$
d \mathrm{X}_{t}^{\varepsilon}=b\left(\varepsilon, \mathrm{X}_{t}^{\varepsilon}\right) d t+\varepsilon \sigma\left(\mathrm{X}_{t}^{\varepsilon}\right) d W_{t}, \quad \text { with } \mathrm{X}_{0}^{\varepsilon}=\mathrm{X}_{0}^{\varepsilon} \in \mathbb{R}^{d},
$$

where $W$ is a $m$-dimensional standard Brownian motion. The main technical result in [17 is a density expansion for $\mathrm{Y}_{t}^{\varepsilon}:=\Pi_{l} \circ \mathrm{X}_{t}^{\varepsilon}$ of the form, for $\mathrm{x}_{0}, \mathrm{y}, T$ fixed,

$$
\begin{equation*}
f^{\varepsilon}(\mathrm{y}, T)=e^{-c_{1} / \varepsilon^{2}} e^{c_{2} / \varepsilon} \varepsilon^{-l}\left(c_{0}+O(\varepsilon)\right) \text { as } \varepsilon \downarrow 0 . \tag{1}
\end{equation*}
$$

One of our main motivations, for [17] and the present paper, comes from the recent work on asset price density expansions by A. Gulisashvili and E.M. Stein: in [28, Theorem 2.1] they consider
the uncorrelated Stein-Stein stochastic volatility model. In the appropriate pricing measure the dynamics ar ${ }^{1}$

$$
d S / S=Z d W^{1}, d Z=(a+b Z) d t+c d W^{2}
$$

with parameters, $a \geq 0, b \leq 0, c>0$, spot-volatility $Z_{0}=\sigma_{0} \geq 0$ and correlation $\rho:=$ $d\left\langle W^{1}, W^{2}\right\rangle / d t$ equal to zero; spot $S_{0}$ can normalized to unit. Their main result is that $S_{T}$, for fixed $T>0$, admits a probability density function $f=f(s)$ such that ${ }^{2}$

$$
\begin{equation*}
f(s)=s^{-B_{1}} e^{B_{2} \sqrt{\log s}}(\log s)^{-\frac{1}{2}}\left(B_{0}+O(\log s)^{-\frac{1}{2}}\right) \text { as } s \uparrow \infty \tag{2}
\end{equation*}
$$

with explicitly computable constants; asymptotic formulae of the implied volatility in the large strike regime are then obtained as (nowadays mechanical; cf. Lee [23] and the references therein) corollaries. Indeed, one ha: $3^{3}$

$$
\begin{align*}
\sigma_{B S}(k, T)^{2} T & =\left(\beta_{1} \sqrt{k}+\beta_{2}+o(1)\right)^{2} \text { as log-strike } k \rightarrow \infty  \tag{3}\\
\beta_{1} & =\beta_{1}\left(B_{1}\right)=\sqrt{2}\left(\sqrt{B_{1}-1}-\sqrt{B_{1}-2}\right) \\
\beta_{2} & =\beta_{2}\left(B_{1}, B_{2}\right)=\frac{B_{2}}{\sqrt{2}}\left(\frac{1}{\sqrt{B_{1}-2}}-\frac{1}{\sqrt{B_{1}-1}}\right)
\end{align*}
$$

The proof of [28, Theorem 2.1] relies on the so-called Hull-White formula which states that - under the crucial assumption of zero correlation - an option price in a stochastic volatility models is effectively a weighted average of Black-Scholes option prices (at different volatiliy levels). The correlated case was left as open problem in [28, Theorem 2.1] and indeed the importance of allowing for correlation in stochastic volatility models is well-documented, e.g. [24, 39]. Evidence from estimation of parametric stochastic volatility models suggests correlation parameter $\rho \approx-0.7$ or $\rho \approx-0.8$ for S\&P 500, for instance; a finding fairly robust across models and time periods [1].

When writing the expansion (2) in terms of $\log$-price $Y=\log S$, it indeed has the form (1) with $y=\log s=1 / \varepsilon^{2}$ and $c_{1}=B_{1}-1, c_{2}=B_{2}$. More generally, we can show from rather general and robust principles that the tail behaviour of $Y_{T} \in \mathbb{R}^{1}$ for fixed $T>0$, subject to a certain scaling with parameter $\theta \in\{1,2\}$ in the full Markovian specification of the model, has the form

$$
\begin{equation*}
f(y, T)=e^{-c_{1} y^{2 / \theta}} e^{c_{2} y^{1 / \theta}} y^{\frac{1}{\theta}-1}\left(c_{0}+O\left(1 / y^{1 / \theta}\right)\right) \text { as } y \uparrow \infty . \tag{4}
\end{equation*}
$$

Again, such an expansion leads immediately to call price and then (Black-Scholes) implied volatility expansions in the large strike regime, cf. [28, 23]; in particular, in the case $\theta=2$ typical for stochastic volatility (see [21] for similar results in the Heston model) the expansion (3) remains valid with $B_{1}=c_{1}+1$ and $c_{2}=B_{2}$. We note that the square-root growth of implied volatility, in terms of log-strike, is actually a very general feature of models with moment explosions, [38, 7] which includes many stochastic volatility models [24, 39, 8].

The main contribution of this paper is to establish validity of (2), equivalently (4) with $\theta=2$, for the correlated Stein-Stein model. Having in mind the typical values o $\rho$ in equity markets, our

[^0]focus is on the case $-1<\rho \leq 0$ (although our analysis could be adapted to positive correlation). The leading order behaviour described by $\beta_{1}=\beta_{1}\left(c_{1}+1\right)$ is well understood; see 38, 7 a and also [18, p40, p265]. The second order behaviour is given by $\beta_{2}=\beta_{2}\left(c_{1}+1, c_{2}\right)$. Further terms in this expansion are in principle possible [23]; in particular, the next term would involve $c_{0}$. Our main observation is that the Stein-Stein model has the scaling properties necessary to transform it into a small noise problem which can then be tackled with the methods of [17]. It should be noted that the Stein-Stein model is hypoelliptic, with region of degeneracy given by $\{(y, z): z=0\}$, and that the $\varepsilon$-rescaled Stein-Stein model is started (as $\varepsilon \rightarrow 0$ ) in the degenerate region. In other words, there is no escape in dealing with the hypoellipticity of the problem $\underline{4}^{4}$

Density expansions of diffusions in the small noise regime, with applications to implied volatility expansions, were recently considered by Y. Osajima 44], based on joint work with S. Kusuoka 36 and old work of Kusuoka-Strook 37. We partially improve on these results. First, as was already mentioned in (17, any expansion of the form (1), or (4), with $c_{2} \neq 0$ is out of reach in these works, the reason being that the Kusuoka-Stroock theory was set up as expansion in $\varepsilon^{2}$ rather than $\varepsilon$. Secondly, in comparison with [44, we do not assume $\mathrm{x}_{0}$ near ( $\left.\mathrm{y}, \cdot\right)$. And finally, in further contrast to (the general results in) [36, 37] we provide a checkable, finite-dimensional criterion that guarantees that the crucial infinite-dimensional non-degeneracy assumption, left as such in 36, 37, is actually satisfied. On the other hand, these authors give explicit formulae for $c_{0}$ which we (presently) do not. Let us also emphasize (cf. corollary 4 below) that the expansion (11) can be used, as a simple consequence of Brownian scaling, towards short time expansion for projected diffusion densities, under global conditions on ( $\mathrm{x}_{0}, \mathrm{y}$ ), of the form

$$
\begin{equation*}
f(\mathrm{y}, t) \sim e^{-\frac{d^{2}\left(\mathrm{x}_{0}, \mathrm{y}\right)}{2 t}} t^{-l / 2} c_{0}\left(\mathrm{x}_{0}, \mathrm{y}\right) \text { as } t \downarrow 0 \tag{5}
\end{equation*}
$$

When $l=d$, and then $\mathrm{y}=\mathrm{x}$, such expansions go back to classical works starting with Molchanov 42 (itself the main reference for the famous SABR paper, 31). The leading order behaviour $2 t \log f(\mathrm{x}, t) \sim-d^{2}\left(\mathrm{x}_{0}, \mathrm{x}\right)$ is due to Varadhan [53]. The case $l<d$, in particular our global condition on $\left(\mathrm{x}_{0}, \mathrm{y}\right)$, appears to be new. That said, expansions of this form have appeared in [52, 31, 44]; the last two references aimed at implied volatility expansions. In the context of a time-homogenuous local volatility models $(l=d=1)$, the expansion (5) holds trivially without any conditions on ( $\mathrm{x}_{0}, \mathrm{y}$ ); the resulting expansion was derived (with explicit constant $c_{0}$ ) in [26]. Subject to mild technical conditions on the diffusion coefficient, they show how to deduce first a call price and then an implied volatility expansion in the short time (to maturity) regime:

$$
\sigma_{B S}(k, t)=|k| / d\left(\mathrm{x}_{0}, k\right)+c\left(\mathrm{x}_{0}, k\right) t+O\left(t^{2}\right) \text { as } t \downarrow 0
$$

where $d\left(\mathrm{x}_{0}, k\right)$ is a point-point distance and $c\left(\mathrm{x}_{0}, k\right)$ is explicitly given. The celebrated Berestycki-Busca-Florent (BBF) formula [14] asserts that $\sigma_{B S}(k, t) \sim|k| / d\left(\mathrm{x}_{0}, k\right)$ as $t \downarrow 0$, is in fact valid in generic stochastic volatilty models, $d\left(\mathrm{x}_{0}, k\right)$ is then understood as point-hyperplane distance. In fact, $|k| / d\left(\mathrm{x}_{0}, k\right)$ arose as initial condition of a non-linear evolution equation for the entire implied volatility surface. As briefly indicated in [14, Sec 6.3] this can be used for a Taylor expansion of $\sigma_{B S}(k, t)$ in $t$. Such expansions have also been discussed, based on heat kernel expansions on Riemannian manifolds by [16, 32, 45], not always in full mathematical rigor. Some mathematical results are given in [44, assuming ellipticity and close-to-the-moneyness $|k| \ll 1$; see also forthcoming work by Ben Arous-Laurence [11]. We suspect that our formula (5), potentially applicable far-from-the-money, will prove useful in this context and shall return to this in future work.

[^1]It should be noted, that the BBF formula alone can be obtained from soft large deviation arguments, cf. [46, Sec. 3.2.1] and the references therein. In a similar spirit, cf. [54, Sec 5, Rmk 2.9], the Varadhan-type formula $2 t \log f(\mathrm{y}, t) \sim-d^{2}\left(\mathrm{x}_{0}, \mathrm{y}\right)$, when $l<d$, could be shown, without any conditions on ( $\mathrm{x}_{0}, \mathrm{y}$ ) by large deviation methods, only relying on the existence of a reasonable density.

As a final note, we recall that the (in general, non-Markovian) $\mathbb{R}^{l}$-valued Itô-process $\left(\mathrm{Y}_{t}: t \geq 0\right)$ admits - subject to some technical assumptions [30, 47] - a Markovian (or Gyöngy) projection. That is, a time-inhomogeneous Markov diffusion ( $\tilde{\mathrm{Y}}_{t}: t \geq 0$ ) with matching time-marginals i.e $\mathrm{Y}_{t}=\tilde{\mathrm{Y}}_{t}$ (in law) for every fixed $t \geq 0$. In a financial context, when $l=1$, this process is known as (Dupire) local volatility model and various authors [14, 16, 32, 11] have used this as an important intermediate step in computing implied volatility in stochastic volatility models. Since all our expansions (small noise, tail, short time ) are relative to such time-marginals they may also be viewed as expansions for the corresponding Markovian projections.

Acknowledgement: JDD and AJ acknowledge (partial resp. full) finanical support from MATHEON. PKF acknowledges partial support from MATHEON and the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement nr. 258237. PKF would like to thank G. Ben Arous for pointing out conceptual similarities in [21, 9 and several discussions thereafter. It is also a pleasure to thank F. Baudoin, J.P. Gauthier, A. Gulisashvili and P. Laurence for their interest and feedback.

## 2 The main result of [17]

Consider a $d$-dimensional diffusion $\left(\mathrm{X}_{t}^{\varepsilon}\right)_{t \geq 0}$ given by the stochastic differential equation

$$
\begin{equation*}
d \mathrm{X}_{t}^{\varepsilon}=b\left(\varepsilon, \mathrm{X}_{t}^{\varepsilon}\right) d t+\varepsilon \sigma\left(\mathrm{X}_{t}^{\varepsilon}\right) d W_{t}, \quad \text { with } \mathrm{X}_{0}^{\varepsilon}=\mathrm{x}_{0}^{\varepsilon} \in \mathbb{R}^{d} \tag{6}
\end{equation*}
$$

and where $W=\left(W^{1}, \ldots, W^{m}\right)$ is an $m$-dimensional Brownian motion. Unless otherwise stated, we assume $b:[0,1) \times \mathbb{R}^{d} \rightarrow \mathbb{R}^{d}, \sigma=\left(\sigma_{1}, \ldots, \sigma_{m}\right): \mathbb{R}^{d} \rightarrow \operatorname{Lin}\left(\mathbb{R}^{m}, \mathbb{R}^{d}\right)$ and $\mathrm{x}_{0}:[0,1) \rightarrow \mathbb{R}^{d}$ to be smooth, bounded with bounded derivatives of all orders. Set $\sigma_{0}=b(0, \cdot)$ and assume that, for every multiindex $\alpha$, the drift vector fields $b(\varepsilon, \cdot)$ converges to $\sigma_{0}$ in the sens $5^{5}$

$$
\begin{equation*}
\partial_{x}^{\alpha} b(\varepsilon, \cdot) \rightarrow \partial_{x}^{\alpha} b(0, \cdot)=\partial_{x}^{\alpha} \sigma_{0}(\cdot) \text { uniformly on compacts as } \varepsilon \downarrow 0 \tag{7}
\end{equation*}
$$

We shall also assume that

$$
\begin{equation*}
\partial_{\varepsilon} b(\varepsilon, \cdot) \rightarrow \partial_{\varepsilon} b(0, \cdot) \quad \text { uniformly on compacts as } \varepsilon \downarrow 0 \tag{8}
\end{equation*}
$$

and

$$
\begin{equation*}
\mathrm{x}_{0}^{\varepsilon}=\mathrm{x}_{0}+\varepsilon \hat{\mathrm{x}}_{0}+o(\varepsilon) \text { as } \varepsilon \downarrow 0 . \tag{9}
\end{equation*}
$$

Theorem 1 (Small noise) Let $\left(\mathrm{X}^{\varepsilon}\right)$ be the solution process to

$$
d \mathrm{X}_{t}^{\varepsilon}=b\left(\varepsilon, \mathrm{X}_{t}^{\varepsilon}\right) d t+\varepsilon \sigma\left(\mathrm{X}_{t}^{\varepsilon}\right) d W_{t}, \quad \text { with } \mathrm{X}_{0}^{\varepsilon}=\mathrm{x}_{0}^{\varepsilon} \in \mathbb{R}^{d}
$$

Assume $b(\varepsilon, \cdot) \rightarrow \sigma_{0}(\cdot)$ in the sense of (7), (8), and $\mathrm{X}_{0}^{\varepsilon} \equiv \mathrm{x}_{0}^{\varepsilon} \rightarrow \mathrm{x}_{0}$ as $\varepsilon \rightarrow 0$ in the sense of (9)). Assume the weak Hörmander condition (H) at $\mathrm{x}_{0} \in \mathbb{R}^{d}$;

$$
\begin{equation*}
\operatorname{span}\left[\sigma_{i}: 1 \leq i \leq m ; \quad\left[\sigma_{j}, \sigma_{k}\right]: 0 \leq j, k \leq m ; \ldots\right]_{\mathrm{x}_{0}}=\mathbb{R}^{d} ; \tag{H}
\end{equation*}
$$

[^2]Fix $\mathrm{y} \in \mathbb{R}^{l}, N_{\mathrm{y}}:=(\mathrm{y}, \cdot)$ and let $\mathcal{K}_{\mathrm{y}}$ be the the space of all $\mathrm{h} \in H$ s.t. the solution to

$$
d \phi_{t}^{\mathrm{h}}=\sigma_{0}\left(\phi_{t}^{\mathrm{h}}\right) d t+\sum_{i=1}^{m} \sigma_{i}\left(\phi_{t}^{\mathrm{h}}\right) d \mathrm{~h}_{t}^{i}, \phi_{0}^{\mathrm{h}}=\mathrm{x}_{0} \in \mathbb{R}^{d}
$$

satisfies $\phi_{T}^{\mathrm{h}} \in N_{\mathrm{y}}$. We assume $\mathcal{K}_{\mathrm{y}}$ to be non-empty ${ }^{6}$ and the energy

$$
\Lambda(\mathrm{y})=\inf \left\{\frac{1}{2}\|\mathrm{~h}\|_{H}^{2}: \mathrm{h} \in \mathcal{K}_{\mathrm{y}}\right\}
$$

to be a smooth function in a neighbourhood of y. Asssume furthermore
(i) there are only finitely many minimizers, i.e. $\mathcal{K}_{\mathrm{y}}^{\mathrm{min}}<\infty$ where

$$
\mathcal{K}_{\mathrm{y}}^{\min }:=\left\{\mathrm{h}_{0} \in \mathcal{K}_{\mathrm{y}}: \frac{1}{2}\left\|\mathrm{~h}_{0}\right\|_{H}^{2}=\Lambda(\mathrm{y})\right\} ;
$$

(ii) non-degeneracy of the so-called deterministic Malliavin covariance matrix at each minimizer - a sufficient condition met in most ("locally elliptic") financial models reads

$$
\forall \mathrm{h}_{0} \in \mathcal{K}_{\mathrm{y}}^{\min }: \exists t \in[0, T]:\left.\operatorname{span}\left[\sigma_{1}, \ldots, \sigma_{m}\right]\right|_{\phi_{t}^{\mathrm{h}_{0}}}=\mathbb{R}^{d}
$$

(iii) $\mathrm{x}_{0}$ is non-focal for $N_{\mathrm{y}}$ in the sense of [17]. (We shall review below how to check this.)

Then, keeping $\mathrm{x}_{0}, y$ and $T>0$ fixed, there exists $c_{0}=c_{0}\left(\mathrm{x}_{0}, \mathrm{y}, T\right)>0$ such that

$$
\mathrm{Y}_{T}^{\varepsilon}=\Pi_{l} \mathrm{X}_{T}^{\varepsilon}=\left(X_{T}^{\varepsilon, 1}, \ldots, X_{T}^{\varepsilon, l}\right), \quad 1 \leq l \leq d
$$

admits a density with expansion

$$
f^{\varepsilon}(\mathrm{y}, T)=e^{-\frac{\Lambda(\mathrm{y})}{\varepsilon^{2}}} e^{\frac{\max \left\{\Lambda^{\prime}(\mathrm{y}) \cdot \hat{\gamma}_{T}\left(\mathrm{~h}_{0}\right): \mathrm{h}_{0} \in \mathcal{K}_{\mathrm{y}}^{\min }\right\}}{\varepsilon}} \varepsilon^{-l}\left(c_{0}+O(\varepsilon)\right) \text { as } \varepsilon \downarrow 0
$$

Here $\hat{\mathrm{Y}}=\hat{\mathrm{Y}}\left(\mathrm{h}_{0}\right)=\left(\hat{Y}^{1}, \ldots, \hat{Y}^{l}\right)$ is the projection, $\hat{\mathrm{Y}}=\Pi_{l} \hat{\mathrm{X}}$, of the solution to the following (ordinary) differential equation

$$
\begin{align*}
d \hat{\mathrm{X}}_{t} & =\left(\partial_{x} b\left(0, \phi_{t}^{\mathrm{h}_{0}}\left(\mathrm{x}_{0}\right)\right)+\partial_{x} \sigma\left(\phi_{t}^{\mathrm{h}_{0}}\left(\mathrm{x}_{0}\right)\right) \dot{\mathrm{h}}_{0}(t)\right) \hat{\mathrm{X}}_{t} d t+\partial_{\varepsilon} b\left(0, \phi_{t}^{\mathrm{h}_{0}}\left(\mathrm{x}_{0}\right)\right) d t  \tag{10}\\
\hat{\mathrm{X}}_{0} & =\hat{\mathrm{x}}_{0}
\end{align*}
$$

Remark 2 (Smoothness of energy) If $\# \mathcal{K}_{\mathrm{y}}^{\mathrm{min}}=1$ smoothness of the energy is actually a (non-trivial) consequence of the present assumptions and hence need not be assumed; 17]. Note also that in our application to tail asymptotics, with $\theta$-scaling, $\theta \in\{1,2\}$ and scalar variable $y$, it follows from scaling that the energy will be a linear resp. quadratic (and hence smooth) function of $y$.

Remark 3 (Localization) The assumptions on the coefficients $b, \sigma$ in theorem 1 (smooth, bounded with bounded derivatives of all orders) are typical in this context (cf. Ben Arous [9, 10] for instance) but rarely met in practical examples from finance. This difficulty can be resolved by a suitable localization. For instance, as detailed in [17], an estimate of the form

$$
\begin{equation*}
\lim _{R \rightarrow \infty} \lim \sup _{\varepsilon \rightarrow 0} \varepsilon^{2} \log \mathbb{P}\left[\tau_{R} \leq T\right]=-\infty \tag{11}
\end{equation*}
$$

with $\tau_{R}:=\inf \left\{t \in[0, T]: \sup _{s \in[0, t]}\left|\mathrm{X}_{s}^{\varepsilon}\right| \geq R\right\}$ will allow to bypass the boundedness assumptions.

[^3]
## 3 Short time and tail asymptotics

The reduction of short time expansions to small noise expansions by Brownian scaling is classical. In the present context, we have the following statement, taken from [17, Sec. 2.1].

Corollary 4 (Short time) Consider $d \mathrm{X}_{t}=b\left(\mathrm{X}_{t}\right) d t+\sigma\left(\mathrm{X}_{t}\right) d W$, started at $\mathrm{X}_{0}=\mathrm{x}_{0} \in \mathbb{R}^{d}$, with $C^{\infty}$-bounded vector fields such that the strong Hörmander condition holds,

$$
\begin{equation*}
\forall \mathrm{x} \in \mathbb{R}^{d}:\left.\operatorname{Lie}\left[\sigma_{1}, \ldots, \sigma_{m}\right]\right|_{\mathrm{x}}=\mathbb{R}^{d} \tag{H1}
\end{equation*}
$$

Fix $\mathrm{y} \in \mathbb{R}^{l}, N_{\mathrm{y}}:=(\mathrm{y}, \cdot)$ and assume (i),(ii),(iii) as in theorem 1. Let $f(t, \cdot)=f(t, \mathrm{y})$ be the density of $\mathrm{Y}_{t}=\left(\mathrm{X}_{t}^{1}, \ldots, \mathrm{X}_{t}^{l}\right)$. Then, for some constant $c_{0}=c_{0}\left(\mathrm{x}_{0}, \mathrm{y}\right)>0$,

$$
\begin{equation*}
f(\mathrm{y}, t) \sim e^{-\frac{d^{2}\left(\mathrm{x}_{0}, \mathrm{y}\right)}{2 t}} t^{-l / 2} c_{0} \text { as } t \downarrow 0 . \tag{12}
\end{equation*}
$$

where $d\left(\mathrm{x}_{0}, \mathrm{y}\right)$ is the sub-Riemannian distance, based on $\left(\sigma_{1}, \ldots, \sigma_{m}\right)$, from the point $\mathrm{x}_{0}$ to the affine subspace $N_{\mathrm{y}}$.

We also have the following application to the tail behaviour of, say, the first component (i.e. $l=1$ here) of a diffusion processes at a fixed time $T$. As we shall see, the scaling assumption below is met in a number of stochastic volatility models.

Corollary 5 (Tail behaviour) Assume $\mathrm{x}_{0}^{\varepsilon} \rightarrow 0 \in \mathbb{R}^{d}$ as $\varepsilon \rightarrow 0$ and some diffusion process $\mathrm{X}^{\varepsilon}$, started at $\mathrm{x}_{0}^{\varepsilon}$, satisfies the assumptions of theorem 1 with $\mathrm{x}_{0}=0$ and $N=(1, \cdot) \subset \mathbb{R} \times \mathbb{R}^{d-1}$; in particular, $\{0\} \times(1, \cdot)$ is assumed to satisfy condition (i),(ii),(iii). Assume also $\theta$-scaling by which we mean the scaling relation

$$
Y_{T}^{\varepsilon} \stackrel{(\text { law })}{=} \varepsilon^{\theta} Y_{T} \text { where } \quad Y \equiv \Pi_{1} \mathrm{X}
$$

for some $\theta \geq 1$. Then the probability density function of $Y_{T}$ has the expansion

$$
\begin{equation*}
f(y)=e^{-c_{1} y^{\frac{2}{\theta}}} e^{c_{2} y^{\frac{1}{\theta}}} y^{\frac{1}{\theta}-1}\left(c_{0}+O\left(1 / y^{1 / \theta}\right)\right) \text { as } y \rightarrow \infty \tag{13}
\end{equation*}
$$

where

$$
\begin{aligned}
& c_{1}=\Lambda(1) \\
& c_{2}=\hat{Y}_{T} \Lambda^{\prime}(1)=\frac{2 \hat{Y}_{T}}{\theta} \Lambda(1)
\end{aligned}
$$

and $c_{0}>0$. In particular, when $\theta=1$ we have a Gaussian tail behaviour of the precise form

$$
f(y)=e^{-\Lambda(1) y^{2}} e^{2 \hat{Y}_{T} \Lambda(1) y}\left(c_{0}+O(1 / y)\right) ;
$$

while $\theta=2$ leads to the exponential tail of the precise form

$$
f(y)=e^{-\Lambda(1) y} e^{\hat{Y}_{T} \Lambda^{\prime}(1) \sqrt{y}} y^{-1 / 2}\left(c_{0}+O(1 / \sqrt{y})\right) .
$$

Proof. Let $f^{\varepsilon}$ denote the density of $Y_{T}^{\varepsilon}$. Since $f\left(\cdot / \varepsilon^{\theta}\right)=\varepsilon^{\theta} f^{\varepsilon}(\cdot)$ we can take $\cdot=1 \in \mathbb{R}^{l}$, with $l=1$, and apply theorem [1 This yields the claimed expansion in the "large space" variable $y=1 / \varepsilon^{\theta}$; it suffices to rephrase the $\varepsilon$-expansion of theorem 1 in terms of $y$. Another observation is that the assumed scaling implies

$$
\Lambda(y)=y^{2 / \theta} \Lambda(1)
$$

and hence $\Lambda^{\prime}(1)=\frac{2}{\theta} \Lambda(1)$. The rest is obvious.

## 4 Computational aspects

We present briefly the mechanics of the actual computations, in the spirit of the Pontryagin maximum principle (e.g. [50), the aim being to find optimal paths which arrive at $N_{\mathrm{a}}=(\mathrm{a}, \cdot)$, i.e. a given "target" manifold. 7 This formalism is justified by assuming non-degeneracy of the so-called deterministic Malliavin covariance matrix $C\left(\mathrm{~h}_{0}\right)$, at each $\mathrm{h}_{0} \in \mathcal{K}_{\mathrm{a}}^{\min } ;$ cf. [52, 17. As pointed out earlier, a sufficient condition met in most ("locally elliptic") financial models reads

$$
\forall \mathrm{h} \in \mathcal{K}_{\mathrm{a}}: \exists t \in[0, T]:\left.\operatorname{span}\left[\sigma_{1}, \ldots, \sigma_{m}\right]\right|_{\phi_{t}^{\mathrm{h}}}=\mathbb{R}^{d}
$$

(This is proved in [17]; for more information on $C\left(\mathrm{~h}_{0}\right)$ see [17] and the references therein; in particular [15, 52]).

- The Hamiltonian. Based on the SDE (6), with diffusion vector fields $\sigma_{1}, \ldots, \sigma_{m}$ and drift vector field $\sigma_{0}$ (in the $\varepsilon \rightarrow 0$ limit) we define the Hamiltonian

$$
\begin{aligned}
\mathcal{H}(\mathrm{x}, \mathrm{p}) & :=\left\langle\mathrm{p}, \sigma_{0}(\mathrm{x})\right\rangle+\frac{1}{2} \sum_{i=1}^{m}\left\langle\mathrm{p}, \sigma_{i}(\mathrm{x})\right\rangle^{2} \\
& =\left\langle\mathrm{p}, \sigma_{0}(\mathrm{x})\right\rangle+\frac{1}{2}\left\langle\mathrm{p},\left(\sigma \sigma^{T}\right)(\mathrm{x}) \mathrm{p}\right\rangle .
\end{aligned}
$$

Remark the driving Brownian motions $W^{1}, \ldots, W^{m}$ were assumed to be independent. Many stochastic models, notably in finance, are written in terms of correlated Brownians, i.e. with a non-trivial correlation matrix $\Omega=\left(\omega^{i, j}: 1 \leq i, j \leq m\right)$, where $d\left\langle W^{i}, W^{j}\right\rangle_{t}=$ $\omega^{i, j} d t$. The Hamiltonian then becomes

$$
\begin{equation*}
\mathcal{H}(\mathrm{x}, \mathrm{p})=\left\langle\mathrm{p}, \sigma_{0}(\mathrm{x})\right\rangle+\frac{1}{2}\left\langle\mathrm{p},\left(\sigma \Omega \sigma^{T}\right)(\mathrm{x}) \mathrm{p}\right\rangle . \tag{14}
\end{equation*}
$$

- The Hamiltonian ODEs. The following system of ordinary differential equations,

$$
\begin{equation*}
\binom{\dot{\mathrm{x}}}{\dot{\mathrm{p}}}=\binom{\partial_{\mathrm{p}} \mathcal{H}(\mathrm{x}(t), \mathrm{p}(t))}{-\partial_{\mathrm{x}} \mathcal{H}(\mathrm{x}(t), \mathrm{p}(t))}, \tag{15}
\end{equation*}
$$

gives rise to a solution flow, denoted by $\mathrm{H}_{t \leftarrow 0}$, so that

$$
\mathrm{H}_{t \leftarrow 0}\left(\mathrm{x}_{0}, \mathrm{p}_{0}\right)
$$

is the unique solution to the above ODE with initial data ( $\mathrm{x}_{0}, \mathrm{p}_{0}$ ). Our standing (regularity) assumption are more than enough to guarantee uniqueness and local ODE existence. As in [15, p.37], the vector field $\left(\partial_{\mathrm{p}} \mathcal{H},-\partial_{\mathrm{x}} \mathcal{H}\right)$ is complete, i.e.one has global existence. It can be usefult to start the flow backwards with time- $T$ terminal data, say $\left(\mathrm{x}_{T}, \mathrm{p}_{T}\right)$; we then write

$$
\mathrm{H}_{t \leftarrow T}\left(\mathrm{x}_{T}, \mathrm{p}_{T}\right)
$$

for the unique solution to (15) with given time- $T$ terminal data. Of course,

$$
\mathrm{H}_{t \leftarrow T}\left(\mathrm{H}_{T \leftarrow 0}\left(\mathrm{x}_{0}, \mathrm{p}_{0}\right)\right)=\mathrm{H}_{t \leftarrow 0}\left(\mathrm{x}_{0}, \mathrm{p}_{0}\right) .
$$

[^4]- Solving the Hamiltonian ODEs as boundary value problem. As before, $N_{\mathrm{a}}=(\mathrm{a}, \cdot)$ is the given "target" manifold; the analysis laid out in 17 requires in a first step to solve the Hamiltonian ODEs (15) with mixed initial -, terminal - and transversality conditions,

$$
\begin{align*}
\mathrm{x}(0) & =\mathrm{x}_{0} \in \mathbb{R}^{d}, \\
\mathrm{x}(T) & =(\mathrm{a}, \cdot) \in \mathbb{R}^{l} \oplus \mathbb{R}^{d-l} \\
\mathrm{p}(T) & =(\cdot, 0) \in \mathbb{R}^{l} \oplus \mathbb{R}^{d-l} \tag{16}
\end{align*}
$$

Note that this is a $2 d$-dimensional system of ordinary differential equations, subject to $d+l+(d-l)=2 d$ conditions. In general, boundary problems for such ODEs may have more than one, exactly one or no solution. In the present setting, there will always be one or more than one solution. After all, we know [17] that there exists at least one minimizing control $h_{0}$ and can be reconstructed via the solution of the Hamiltonian ODEs, as explained in the following step.

- Finding the minimizing controls. The Hamiltonian ODEs, as boundary value problem, are effectively first order conditions (for minimality) and thus yield candidates for the minimizing control $h_{0}=h_{0}(\cdot)$, given by

$$
\dot{\mathrm{h}}_{0}=\left(\begin{array}{c}
\left\langle\sigma_{1}(\mathrm{x}(\cdot)), \mathrm{p}(\cdot)\right\rangle  \tag{17}\\
\ldots \\
\left\langle\sigma_{m}(\mathrm{x}(\cdot)), \mathrm{p}(\cdot)\right\rangle
\end{array}\right) .
$$

Each such candidate is indeed admissible in the sense $h_{0} \in \mathcal{K}_{a}$ but may fail to be a minimizer. We thus compute the energy $\left\|\mathrm{h}_{0}\right\|_{H}^{2}$ for each candidate and identify those (" $\mathrm{h}_{0} \in \mathcal{K}_{\mathrm{a}}^{\min }$ ") with minimal energy. The procedure via Hamiltonian flows also yields a unique $\mathrm{p}_{0}=\mathrm{p}_{0}\left(\mathrm{~h}_{0}\right)$.

- Checking non-focality. By definition [17], $\mathrm{x}_{0}$ is non-focal for $N=(\mathrm{a}, \cdot)$ along $\mathrm{h}_{0} \in \mathcal{K}_{\mathrm{a}}^{\min }$ in the sense that, with $((\mathrm{a}, \cdot),(\cdot, 0)) \ni\left(\mathrm{x}_{T}, \mathrm{p}_{T}\right):=\mathrm{H}_{T \leftarrow 0}\left(\mathrm{x}_{0}, \mathrm{p}_{0}\left(\mathrm{~h}_{0}\right)\right) \in \mathcal{T}^{*} \mathbb{R}^{d}$,

$$
\left.\partial_{(\mathfrak{z}, \mathfrak{q})}\right|_{(\mathfrak{z}, \mathfrak{q})=(0,0)} \pi \mathrm{H}_{0 \leftarrow T}\left(\mathrm{x}_{T}+\binom{0}{\mathfrak{z}}, \mathrm{p}_{T}+(\mathfrak{q}, 0)\right)
$$

is non-degenerate (as $d \times d$ matrix; here we think of $(\mathfrak{z}, \mathfrak{q}) \in \mathbb{R}^{d-l} \times \mathbb{R}^{l} \cong \mathbb{R}^{d}$ and recall that $\pi$ denotes the projection from $\mathcal{T}^{*} \mathbb{R}^{d}$ onto $\mathbb{R}^{d}$; in coordinates $\left.\pi(\mathrm{x}, \mathrm{p})=\mathrm{x}\right)$. Note that in the point-point setting, $\mathrm{x}_{T}=\mathrm{x}$ for fixed x , only perturbations of the arrival "velocity" $\mathrm{p}_{T}$ - without any restrictions of transversality type - are considered. Non-degeneracy of the resulting map should then be called non-conjugacy (between two points; here: $\mathrm{x}_{T}$ and $\mathrm{x}_{0}$ ). In the Riemannian setting this is consistent with the usual meaning of non-conjugacy; after identifying tangent- and cotangent-space $\left.\partial_{\mathfrak{q}}\right|_{\mathfrak{q}=0} \pi \mathrm{H}_{0 \leftarrow T}$ is precisely the differential of the exponential map.

- The explicit marginal density expansion. We then have

$$
f^{\varepsilon}(\mathrm{a}, T)=e^{-c_{1} / \varepsilon^{2}} e^{c_{2} / \varepsilon} \varepsilon^{-l}\left(c_{0}+O(\varepsilon)\right) \text { as } \varepsilon \downarrow 0
$$

with $c_{1}=\Lambda(\mathrm{a})$. The second-order exponential constant $c_{2}$ then requires the solution of a finitely many $\left(\# \mathcal{K}_{\mathrm{a}}^{\min }<\infty\right)$ auxilary ODEs, cf. theorem 1 At last we set $\mathrm{a}=\mathrm{y} \in \mathbb{R}^{l}$, in context of small noise and short time expansions, and $\mathrm{a}=1 \in \mathbb{R}^{l}$, with $l=1$ for tail expansions (in this case, the $y$-dependence here is hidden in $\varepsilon$ ).

## 5 Application to asset price models

### 5.1 Black-Scholes

The Black-Scholes (BS) model, written in terms of log-price is an example where the above theorem is applicable with $\theta=1$. Indeed, $Y:=\log S$ satisfies, with fixed Black-Scholes volatility $\sigma>0$

$$
d Y_{t}=-\frac{\sigma^{2}}{2} d t+\sigma d W_{t}, \quad Y_{0}=y_{0}=\log S_{0}
$$

Of course, $Y_{t} \sim N\left(y_{0}-\sigma^{2} t / 2, \sigma^{2} t\right)$ and the explicit Gaussian density

$$
f_{\mathrm{BS}}(t, y)=\frac{1}{\sqrt{2 \pi \sigma^{2} t}} \exp \left\{-\frac{\left(y-\left(y_{0}-\sigma^{2} t / 2\right)\right)^{2}}{2 \sigma^{2} t}\right\}
$$

immediately yields short time resp. tail expansions,

$$
\begin{align*}
& f_{\mathrm{BS}}(t, y) \sim(\text { const }) t^{-1 / 2} \exp \left(-\frac{\left(\frac{y-y_{0}}{\sigma}\right)^{2}}{2 t}\right) \text { as } t \downarrow 0 ; \text { any } y \in \mathbb{R}  \tag{18}\\
& f_{\mathrm{BS}}(T, y) \sim(\text { const }) \exp \left(-\frac{1}{2 \sigma^{2} T} y^{2}\right) \exp \left(\frac{y_{0}-\sigma^{2} T / 2}{2 \sigma^{2} T} y\right) \text { as } y \rightarrow \infty ; \text { any } T>0 . \tag{19}
\end{align*}
$$

We derive now both expansions from general theory, i.e. with aid of corollary 4 resp 5. The short time limit corresponds to a flat Riemannian situation, in particular the cutlocus is empty, which is enough to guarantee (ND); the remaining computations to derive (18) from corollary 4 are left to the reader and we focus on the (more interesting) case of tail asymptotics. Corollary 5 applies with $\theta=1$, and (rescaled) starting point $\varepsilon y_{0} \rightarrow 0$. Condition (ND) needs to be checked; the relevant Hamiltonian is

$$
\mathcal{H}(y, p)=\frac{\sigma^{2} p^{2}}{2}, \quad \text { for all }(y, p) \in \mathbb{R}^{2}
$$

and the Hamiltonian ODEs are

$$
\dot{y}_{t}=\sigma^{2} p_{t}, \quad \dot{p}_{t}=0,
$$

with boundary conditions $y_{0}=0$ and $y_{T}=1=$ : $a$. Since $p_{t}$ is constant, we obtain $p_{t} \equiv p_{0}=\frac{a}{\sigma^{2} T}$, and $y_{t}=a \frac{t}{T}$. In particular, $\left.\partial_{p} y_{T}\right|_{p_{0}}=\sigma^{2} T>0$, and hence invertible, for $T, \sigma>0$. En passant, we also deduce the optimal control $h_{0}(t)=\sigma p_{0}$, and get the correct leading order factor

$$
c_{1}:=\frac{1}{2}\left\|h_{0}\right\|^{2}=\frac{1}{2} \int_{0}^{T} h_{0}(t)^{2} d t=\frac{1}{2 \sigma^{2} T} .
$$

With the hint $\hat{Y}_{t}=y_{0}+\left(-\frac{\sigma^{2}}{2}\right) t$ we leave it to the reader to verify that $c_{2}=\left(y_{0}-\sigma^{2} T / 2\right) /\left(2 \sigma^{2} T\right)$. Frequently, one chooses $y_{0}=0$ in this context (which amounts to normalize spot price to unit).

### 5.2 The Stein-Stein model

For given parameters, $a \geq 0, b<0, c>0, \sigma_{0} \geq 0, \rho=d\left\langle W^{1}, W^{2}\right\rangle / d t$, the Stein-Stein model expresses $\log$-price $Y$, under the forward measure, via

$$
\begin{align*}
d Y & =-\frac{1}{2} Z^{2} d t+Z d W^{1}, Y(0)=y_{0}=0  \tag{20}\\
d Z & =(a+b Z) d t+c d W^{2}, \quad Z(0)=\sigma_{0}>0 .
\end{align*}
$$

We will be interested in the behaviour, and in particular the tail-behaviour, of the probability density function of $Y_{T}$. In fact, there is no loss of generality to consider $T=1$. Applying Brownian scaling, it is a straight-forward computation to see that the pair $(\tilde{Y}, \tilde{Z})$ given by

$$
\tilde{Y}(t):=Y(t T), \quad \tilde{Z}(t):=Z(t T) T^{1 / 2}
$$

satisfies the same parametric SDE form as Stein-Stein, but with the following parameter substitutions

$$
a \leftarrow \tilde{a} \equiv a T^{3 / 2}, b \leftarrow \tilde{b} \equiv b T, c \leftarrow \tilde{c} \equiv c T, \sigma_{0} \leftarrow \tilde{\sigma}_{0} \equiv \sigma_{0} T^{1 / 2}
$$

In particular then, $Y_{T}=Y_{T}\left(a, b, c, \sigma_{0}, \rho\right)$ has the same law as $Y_{1}\left(\tilde{a}, \tilde{b}, \tilde{c}, \tilde{\sigma}_{0}, \rho\right)$.

### 5.2.1 The case of zero-correlation

For the moment, we shall follow [28] in assuming the Brownians to be uncorrelated,

$$
d\left\langle W^{1}, W^{2}\right\rangle_{t}=\rho d t \text { with } \rho=0
$$

Recall their main result, a density expansion for $Y_{T}$ of the form

$$
\begin{equation*}
(*): f(y)=e^{-c_{1} y} e^{c_{2} y^{1 / 2}} y^{-1 / 2}\left(c_{3}+O\left(y^{-1 / 2}\right)\right) \text { as } y \rightarrow \infty . \tag{21}
\end{equation*}
$$

Scaling: Setting

$$
Y_{\varepsilon}:=\varepsilon^{2} Y, Z_{\varepsilon}:=\varepsilon Z
$$

yields the small noise problem

$$
\begin{align*}
& d Y_{\varepsilon}=-\frac{1}{2} Z_{\varepsilon}^{2} d t+Z_{\varepsilon} \varepsilon d W^{1}, Y_{\varepsilon}(0)=0=: y_{0} \forall \varepsilon>0  \tag{22}\\
& d Z_{\varepsilon}=\left(a \varepsilon+b Z_{\varepsilon}\right) d t+c \varepsilon d W^{2}, \quad Z_{\varepsilon}(0)=\varepsilon \sigma_{0} \rightarrow 0=: z_{0} \text { as } \varepsilon \downarrow 0
\end{align*}
$$

Our corollary [5, assuming its application to be justified, then gives the correct expansion (21), namely

$$
f(y)=e^{-c_{1} y} e^{c_{2} y^{1 / 2}} y^{-1 / 2}\left(c_{3}+O\left(y^{-1 / 2}\right)\right)
$$

and also identifies the constants $c_{1}=\Lambda(1), c_{2}=\hat{Y}_{T} \Lambda^{\prime}(1)$. (The leading order constant $c_{1}$ is in agreement with both [28] and [18, p40].)

Remark 6 Corollary 5 relies on an application of theorem 1 to (22); let us note straight away that the coefficients here are smooth but unbounded. With a view towards the earlier remark on localization, and in particular (11), we note here that, due to the particular structure of the SDE, it suffices to localize such as to make $\sigma$ bounded; e.g. by stopping it upon leaving a big ball of radius $R$. This amounts to, cf. (11), to shows that

$$
\lim _{R \rightarrow \infty} \lim \sup _{\varepsilon \rightarrow 0} \varepsilon^{2} \log \mathbb{P}\left[\left|\sigma_{\varepsilon}\right|_{\infty ;[0, T]} \geq R\right]=-\infty
$$

But since $\mathbb{P}\left[\left|\sigma_{\varepsilon}\right|_{\infty ;[0, T]} \geq R\right]=\mathbb{P}\left[|\sigma|_{\infty ;[0, T]} \geq R / \varepsilon\right]$ and $\sigma$ is a Gaussian process, this is an immediate consequence of Fernique's estimate.

We postpone the justification that we may indeed apply corollary 5 (which involves an analysis of the Hamiltonian ODEs) and proceed in showing how further qualitative information about the expansion can be obtained without much computations.
Some information on $c_{1}$ : According to theorem 1

$$
c_{1}:=\Lambda(1)=\inf \left\{\frac{1}{2}\|\mathrm{~h}\|_{H}^{2}: \phi_{0}^{\mathrm{h}}=(0,0), \phi_{T}^{\mathrm{h}} \in(1, \cdot)\right\}
$$

where $d \phi_{t}^{\mathrm{h}, 1}=-\frac{1}{2}\left|\phi_{t}^{\mathrm{h}, 2}\right|^{2} d t+\phi_{t}^{\mathrm{h}, 2} d h^{1}, d \phi_{t}^{\mathrm{h}, 2}=b \phi_{t}^{\mathrm{h}, 2} d t+c d h^{2}$. If then follows a priori that

$$
c_{1}=c_{1}(b, c ; T) \text { but not on } a, \sigma_{0} .
$$

The same is true for $\mathrm{h}_{0}=: \mathrm{h}^{*}=:\left(h^{*, 1}, h^{*, 2}\right)$ and $\phi^{*}:=\phi^{\mathrm{h}_{0}}$ of course.
Some information on $c_{2}$ : First, $\Lambda^{\prime}(1)=c_{1}$ also only depends on the parameters $b, c, T$ (but not on $\left.a, \sigma_{0}\right)$. It remains to analyze the factor $\hat{Y}_{T}$ where $\left(\hat{Y}_{t}, \hat{Z}_{t}: t \geq 0\right)$ solves the ODE

$$
\begin{aligned}
& d \hat{Y}_{t}=\left(-\phi_{t}^{*, 2}+h_{t}^{*, 1}\right) \hat{Z}_{t} d t, \quad \hat{Y}_{0}=0 \\
& d \hat{Z}_{t}=b \hat{Z}_{t} d t+a d t, \quad \hat{Z}_{0}=\sigma_{0}
\end{aligned}
$$

Since $\hat{Z}_{t}=\sigma_{0} e^{b t}+a \int_{0}^{t} e^{b(t-s)} d s$ it follows that $\hat{Z}_{T}$ is linear in $\sigma_{0}, a$ with coefficients depending on $b$ and $T$. Furthermore, noting that

$$
\hat{Y}_{T}=\int_{0}^{T}\left(-\phi_{t}^{*, 2}+h_{t}^{*, 1}\right) \hat{Z}_{t} d t
$$

a similar statement is true for $\hat{Y}_{T}$ and then $c_{2}=\Lambda^{\prime}(1) \times \hat{Y}_{T}^{1}$. Namely, for constants $C_{i}=$ $C_{i}(b, c ; T)$

$$
c_{2}=C_{1}(b, c ; T) \sigma_{0}+C_{2}(b, c ; T) a
$$

It is interesting to compare this with the Heston result [21] where the constant $c_{2}$ also depends linearly on spot-vol $\sigma_{0}=\sqrt{v_{0}}$.
Solving the Hamiltonian ODEs and computing $c_{1}$ After replacing $\varepsilon d W$ by a control $d \mathrm{~h}$, and taking $\varepsilon \downarrow 0$ elsewhere in (22), we have to consider the controlled ordinary differential equation

$$
\begin{align*}
d y & =-\frac{1}{2} z^{2} d t+z d h^{1}, y_{0}=0  \tag{23}\\
d z & =b z d t+c d h^{2}, \quad z_{0}=0
\end{align*}
$$

minimizing the energy, $\frac{1}{2} \int_{0}^{T}\left|\dot{\mathrm{~h}}_{t}\right|^{2} d t$ subject to $y_{T}=\mathrm{a} \equiv 1>0$.
According to general theory, we now write out the Hamiltonian associated to (23),

$$
\begin{align*}
& \mathcal{H}\left(\binom{y}{z} ;(p, q)\right)  \tag{24}\\
= & \binom{-\frac{1}{2} z^{2}}{b z} \cdot\binom{p}{q}+\frac{1}{2}\left|\binom{z}{0} \cdot\binom{p}{q}\right|^{2}+\frac{1}{2}\left|\binom{0}{c} \cdot\binom{p}{q}\right|^{2} \\
= & -\frac{1}{2} z^{2} p+b z q+\frac{1}{2}\left(z^{2} p^{2}+c^{2} q^{2}\right) .
\end{align*}
$$

The Hamiltonian ODEs then become

$$
\begin{align*}
\binom{\dot{y}_{t}}{\dot{z}_{t}} & =\binom{z_{t}^{2}\left(p_{t}-\frac{1}{2}\right)}{b z_{t}+c^{2} q_{t}}  \tag{25}\\
\binom{\dot{p}_{t}}{\dot{q}_{t}} & =\binom{0}{p_{t} z_{t}\left(1-p_{t}\right)-b q_{t}} .
\end{align*}
$$

Trivially, $p_{t} \equiv p_{0}$ which we shall denote by $p$ from here on. As it turns out there is a simple expression for the energy. Although we shall ultimately take $\mathrm{a} \equiv 1$ it is convenient to carry out the following analysis for general $\mathrm{a}>0$.

Lemma 7 For any $\mathrm{h}_{0} \in \mathcal{K}_{\mathrm{a}}^{\min }$, and in fact any $\mathrm{h}_{0}$ given by (17), i.e.

$$
\begin{equation*}
\dot{\mathrm{h}}_{0}(t)=\binom{p z_{t}}{q_{t} c} \tag{26}
\end{equation*}
$$

where $(y, z ; p, q)$ satisfies (25), subject to boundary conditions $\left(y_{0}, z_{0}\right)=(0,0)$ and $y_{T}=\mathrm{a}, q_{T}=0$, we have

$$
\Lambda(\mathrm{a})=\frac{1}{2} \int_{0}^{T}\left|\dot{\mathrm{~h}}_{0}(t)\right|^{2} d t=p \mathrm{a}
$$

In particular, we see that

$$
p \geq 0
$$

Remark 8 In fact, linearity in a of (34) also follows immediately from the fact that the SteinStein model satisfies $\theta$-scaling with $\theta=2$ in the sense of corollary 5. Indeed, it was seen in the proof of that corollary that the rate function $\Lambda$ (a) scales like $\mathrm{a}^{2 / \theta}=\mathrm{a}$. This already implies that $p$ does not depend on a . This is also consistent with the principle $\partial_{\mathrm{a}} \Lambda(\mathrm{a})=p_{T}$ pointed out in 17.

Proof. We give an elegant argument based on the Hamiltonian ODEs. The idea is to express $\left|\dot{\mathrm{h}}_{0}(t)\right|^{2}$ as a time-derivative which then allows for immediate integration over $t \in[0, T]$. Indeed,

$$
\begin{aligned}
\left|\dot{\mathrm{h}}_{0}(t)\right|^{2} & =p^{2} z_{t}^{2}+c^{2} q_{t}^{2} \\
& =p^{2} z_{t}^{2}+\partial_{t}\left(z_{t} q_{t}\right)-z_{t}^{2}\left(p^{2}-p\right) \\
& =2 p z_{t}^{2}(p-1 / 2)+\partial_{t}\left(z_{t} q_{t}\right) \\
& =2 p \dot{y_{t}}+\partial_{t}\left(z_{t} q_{t}\right)
\end{aligned}
$$

where we used the ODEs for $z, q$ as given in (25). It follows that

$$
\int_{0}^{T}\left|\dot{\mathrm{~h}}_{0}(t)\right|^{2} d t=2 p\left(y_{T}-y_{0}\right)+\left(z_{T} q_{T}-z_{0} q_{0}\right)
$$

and we conclude with the initial/terminal/transversality conditions $y_{0}=z_{0}=0, y_{T}=\mathrm{a}$ and $q_{T}=0$.
Lemma 9 (Partial Hamiltonian Flow) Consider (25) as initial value problem, with initial data $\left(y_{0}, z_{0}\right)=(0,0)$ and $\left(p, q_{0}\right)$. Assum』 ${ }^{8}$

$$
\begin{equation*}
\chi_{p}^{2}:=c^{2} p(p-1)-b^{2} \geq 0 \tag{27}
\end{equation*}
$$

[^5]Then the explicit solution is given by

$$
\begin{align*}
& y_{t}=\frac{q_{0}^{2} c^{4}\left(2 p_{0}-1\right)}{8 \chi_{p}^{3}}\left(2 \chi_{p} t-\sin \left(2 \chi_{p} t\right)\right)  \tag{28}\\
& z_{t}=\frac{q_{0} c^{2}}{\chi_{p}} \sin \left(\chi_{p} t\right) \\
& p_{t} \equiv p \\
& q_{t}=q_{0}\left(\cos \left(\chi_{p} t\right)-\frac{b}{\chi_{p}} \sin \left(\chi_{p} t\right)\right)
\end{align*}
$$

Remark 10 The given solutions remain valid when $\chi_{p}^{2}<0$; it suffices to consider $\chi_{p}$ as purely imaginary; then, if desired, rewrite as $\cos \left(\chi_{p} t\right)=\cosh \left(\left|\chi_{p}\right| t\right)$ etc. Below, we shall solve (25) as boundary value problem, subject to $\left(y_{0}, z_{0}\right)=(0,0), y_{T}=\mathrm{a}>0$ and $q_{T}=0$; we shall see then that (27) is always satisfied and in fact $\chi_{p}^{2}>0$.

Proof. Let us first remark that the path $\left(p_{t}\right)_{t \geq 0}$ is constant, $p_{t}=p$ for all $t \in[0, T]$. From the Hamiltonian ODEs, the couple $\left(z_{t}, q_{t}\right)_{t \geq 0}$ solves a linear ODE in $\mathbb{R}^{2}$, so that the solution must be a linear function of $\left(z_{0}, q_{0}\right)=\left(0, q_{0}\right)$. Indeed, a simple computation gives

$$
q_{t}=q_{0}\left(\cos \left(\chi_{p} t\right)-\frac{b}{\chi_{p}} \sin \left(\chi_{p} t\right)\right) \quad \text { and } \quad z_{t}=\frac{q_{0} c^{2}}{\chi_{p}} \sin \left(\chi_{p} t\right)
$$

Elementary integration $\left(" 2 \int_{0}^{t} \sin ^{2}=t-\cos \sin t "\right)$ then gives $\left(y_{t}\right)_{t \geq 0}$ by direct integration; indeed

$$
y_{t}=\left(p-\frac{1}{2}\right) \int_{0}^{t} z_{s}^{2} d s=\frac{q_{0}^{2} c^{4}(2 p-1)}{8 \chi_{p}^{3}}\left(2 \chi_{p} t-\sin \left(2 \chi_{p} t\right)\right) .
$$

This proves the lemma.
For the next proposition we recall the standing assumptions $T>0, b \leq 0$ (which models mean-reversion) and $\mathrm{a}>0$.

Proposition 11 The ensemble of solutions to the Hamilton ODEs as boundary value problem

$$
\left(y_{0}, z_{0}\right)=(0,0) \text { and } y_{T}=\mathrm{a}, q_{T}=0
$$

with $\mathrm{a}=1>0$ are characterized by inserting, for any $k \in\{1,2, \ldots\}$ and any choice of sign in (30) below,

$$
\begin{align*}
p & =p_{k}=\frac{1}{2}\left(1+\sqrt{1+\frac{4 b^{2}}{c^{2}}+\frac{4 r_{k}^{2}}{c^{2} T^{2}}}\right),  \tag{29}\\
q_{0, k}^{ \pm} & = \pm \frac{2}{c^{2}} \sqrt{\frac{2 r_{k}^{3} \mathrm{a}}{\left(2 p_{0, k}^{+}-1\right) T^{3}\left(2 r_{k}-\sin \left(2 r_{k}\right)\right)}} \tag{30}
\end{align*}
$$

in (28). Here $\left\{r_{k}: k=1,2, \ldots\right\}$ denotes the set of (increasing) strictly positive roots to

$$
r \cos (r)-b T \sin (r)=0
$$

Remark 12 As the proof will show, $p$ as given in (29) is the unique positive root to

$$
c^{2} p(p-1)-b^{2}=\left(\frac{r_{0, k}}{T}\right)^{2}
$$

in particular, assumption (27) in the previous lemma is met.
Proof. By assumption and (28),

$$
\begin{equation*}
0=q_{T}=q_{0}\left(\cos \left(\chi_{p} T\right)-\frac{b}{\chi_{p}} \sin \left(\chi_{p} T\right)\right) \tag{31}
\end{equation*}
$$

At this stage, $\chi_{p}$ could be a complex number (when $\chi_{p}^{2}<0$ ). Let us note straight away that we must have $q_{0} \neq 0$ for otherwise $\left(y_{t}\right)_{t \geq 0}$ - which depends linearly on $q_{0}$ as is seen explicitly in (28) - would be identically equal to zero in contradiction with $y_{T}=\mathrm{a}>0$. Let us also note that $\chi_{p} \neq 0$ for otherwise (31), which has a removable singularity at $\chi_{p}=0$, leads to the contradiction $0=1-b T$. (Recall $b \leq 0, T>0$.) But then $r:=\chi_{p} T$ is a root, i.e. maps to zero, under the map

$$
\begin{equation*}
r \in \mathbb{C} \mapsto r \cos r-b T \sin r=r\left(\cos r-\frac{b T}{r} \sin r\right) \tag{32}
\end{equation*}
$$

A complex analysis lemma [28, Lemma 4] asserts that this map, provided

$$
\begin{equation*}
-b T \geq 0 \tag{33}
\end{equation*}
$$

has only real roots; it follows that $\chi_{p}$ is real and so $\chi_{p}^{2} \geq 0$; actually $\chi_{p}^{2}>0$, since we already noted that $\chi_{p} \neq 0$. Note that (31), and in fact all further expressions involving $\chi_{p}$, are unchanged upon changing sign of $\chi_{p}$, we shall agree to take $\chi_{p}>0$ as the positive square-root of $\chi_{p}^{2}$. In particular, (31) is equivalent to the existence of $\chi_{p}>0$ such that

$$
\chi_{p} T \cos \left(\chi_{p} T\right)-b T \sin \left(\chi_{p} T\right)=0
$$

It follows that $\chi_{p} T \in\left\{r_{k}: k=0,1,2, \ldots\right\}$, the set of zeros of (32) written in increasing order. We deduce that, for each $k=0,1,2, \ldots$ there is a choice of $p$ arising from

$$
\chi_{p}^{2}=c^{2} p(p-1)-b^{2}=\left(\frac{r_{k}}{T}\right)^{2}
$$

For each $k$, there is a negative solution, say $p=p_{k}^{-}<0$ which we may ignore thanks to lemma 7 and a positive solution, namely

$$
p=p_{k}^{+}=\frac{1}{2}\left(1+\sqrt{1+\frac{4 b^{2}}{c^{2}}+\frac{4 r_{k}^{2}}{c^{2} T^{2}}}\right)>1
$$

We now exploit $y_{T}=\mathrm{a}$. From the explicit expression of $y_{t}$ given in (28) we get

$$
\begin{aligned}
\mathrm{a} & =y_{T}=\frac{q_{0}^{2} c^{4}(2 p-1)}{8 \chi_{p}^{3}}\left(2 \chi_{p} T-\sin \left(2 \chi_{p} T\right)\right) \\
& =\frac{q_{0}^{2} c^{4}(2 p-1) T^{3}}{8 r_{k}^{3}}\left(2 r_{k}-\sin \left(2 r_{k}\right)\right)
\end{aligned}
$$

and thus

$$
q_{0}^{2}=\frac{8 r_{k}^{3}}{c^{4}(2 p-1) T^{3}\left(2 r_{k}-\sin \left(2 r_{k}\right)\right)} \mathrm{a}
$$

It follows that, for each $k \in\{1,2, \ldots\}$, we can take

$$
\begin{aligned}
p & =p_{k}^{+}=\frac{1}{2}\left(1+\sqrt{1+\frac{4 b^{2}}{c^{2}}+\frac{4}{c^{2}}\left(\frac{r_{k}}{T}\right)^{2}}\right) \\
q_{0} & =q_{0, k}^{ \pm}
\end{aligned}= \pm \frac{2}{c^{2}} \sqrt{\frac{2 r_{k}^{3} \mathrm{a}}{\left(2 p_{k}^{+}-1\right) T^{3}\left(2 r_{k}-\sin \left(2 r_{k}\right)\right)}}
$$

and any such choice in (28) leads to a solution of the boundary value problem.

So far, we have for each $k \in\{1,2, \ldots\}$ two choices of $\left(p, q_{0}\right)$, depending on the sign in (30) so that the resulting Hamiltonian ODE solutions, started from $\left(y_{0}, z_{0}\right)=(0,0)$ and $\left(p, q_{0}\right)$, describe all possible solutions of the boundary value problem given by the Hamiltonian ODEs with mixed initial/terminal data

$$
\left(y_{0}, z_{0}\right)=(0,0) \text { and } y_{T}=\mathrm{a}, q_{T}=0 .
$$

It remains to see which choice (or choices) lead to minimizing controls; i.e. $\mathrm{h}_{0} \in \mathcal{K}_{\mathrm{a}}^{\min }$. But this is easy since we know from lemma 7 that, for any $p \in\left\{p_{k}^{+}: k=1,2, \ldots\right\}$,

$$
\frac{1}{2} \int_{0}^{T}\left|\dot{\mathrm{~h}}_{0}(t)\right|^{2} d t=p \mathrm{a}
$$

Since $p_{k}^{+}$is plainly (strictly) increasing in $k \in\{1,2, \ldots\}$, we see that the energy is minimal if and only if $p=p_{1}^{+}$. On the other hand, we are left with two choices for $q_{0}$, namely $q_{0,1}^{+}$and $q_{0,1}^{-}$. Using (26) we then see that there are two minimizing controls,

$$
\mathcal{K}_{\mathrm{a}}^{\min }=\left\{\mathrm{h}_{0}^{+}, \mathrm{h}_{0}^{-}\right\}
$$

given by

$$
\dot{\mathrm{h}}_{0}^{ \pm}(t)=\binom{p \frac{q_{0} c^{2}}{\chi_{p}} \sin \left(\chi_{p} t\right)}{c q_{0}\left(\cos \left(\chi_{p} t\right)-\frac{b}{\chi_{p}} \sin \left(\chi_{p} t\right)\right)} \quad \text { with }\left(p, q_{0}\right) \leftarrow\left(p_{1}^{+}, q_{0,1}^{+}\right) \text {resp. }\left(p_{1}^{+}, q_{0,1}^{-}\right)
$$

Of course, $\mathrm{h}_{0}^{ \pm}$stands for $\mathrm{h}_{0}^{+}$resp. $\mathrm{h}_{0}^{-}$depending on the chosen substitution above. In $(y, z)$ coordinates, note that both $\mathrm{h}_{0}^{+}$and $\mathrm{h}_{0}^{-}$have identical $y$-components; their $z$-components only differ by a flipped sign due to $q_{0,1}^{-}=-q_{0,1}^{+}$. (This reflects a fundamental symmetry in our problem which is in fact invariant under $(y, z) \mapsto(y,-z))$. We summarize our finds in stating that

$$
\begin{equation*}
\Lambda(\mathrm{a})=\frac{1}{2}\left\|\mathrm{~h}_{0}^{+}\right\|_{H}^{2}=\frac{1}{2}\left\|\mathrm{~h}_{0}^{-}\right\|_{H}^{2}=p_{1}^{+} \mathrm{a} \tag{34}
\end{equation*}
$$

and upon taking $\mathrm{a}=1$ we have computed the leading order constant

$$
c_{1}=\Lambda(1)=p_{1}^{+}=\frac{1}{2}\left(1+\sqrt{1+\frac{4 b^{2}}{c^{2}}+\frac{4}{c^{2}}\left(\frac{r_{1}}{T}\right)^{2}}\right)
$$

where we recall that $r_{1}$ is the first strictly positive root of the equation $r \cos (r)-b T \sin (r)=0$.

Computing $c_{2}$ According to general theory, cf. equation (10), we need to compute certain ODEs for each minimizer, $\mathrm{h}_{0}^{+}=\left(h_{0, .}^{+, 1}, h_{0, .}^{+, 2}\right)$ resp. $\mathrm{h}_{0}^{-}=\left(h_{0, .}^{-,,}, h_{0, \cdot}^{-, 2}\right)$, exhibited in the previous section. For ease of notation we shall write $\left(p, q_{0}^{ \pm}\right)$instead of $\left(p_{1}^{+}, q_{0,1}^{+}\right)$resp. $\left(p_{1}^{+}, q_{0,1}^{-}\right)$in this section. Related to equation (22) we then have to consider the following ODE along $\mathrm{h}_{0}^{+}$(and then along $\mathrm{h}_{0}^{-}$)

$$
\begin{aligned}
\frac{d}{d t}\binom{\hat{Y}_{t}}{\hat{Z}_{t}^{2}} & =\left\{\left(\begin{array}{cc}
0 & -z_{t}^{+} \\
0 & b
\end{array}\right)+\left(\begin{array}{ll}
0 & 1 \\
0 & 0
\end{array}\right) \dot{h}_{0, t}^{+, 1}\right\}\binom{\hat{Y}_{t}}{\hat{Z}_{t}^{2}}+\binom{0}{a} \\
& =\left(\begin{array}{cc}
0 & (p-1) z_{t}^{+} \\
0 & 0
\end{array}\right)\binom{\hat{Y}_{t}}{\hat{Z}_{t}^{2}}+\binom{0}{a} \\
\text { with }\binom{\hat{Y}_{0}}{\hat{Z}_{0}^{2}} & =\binom{0}{\sigma_{0}} .
\end{aligned}
$$

Here, we used the fact that $\dot{h}_{0}^{+, 1}=p z_{t}^{+}, z_{t}^{+}$indicates the chosen sign of $q_{0,1}$ upon which it depends, cf. (30). The ODE along $\mathrm{h}_{0}^{-}$for $\hat{Y}=\hat{Y}^{-}$is similar, with $z_{t}^{+}, \dot{h}_{0, t}^{+, 1}$ replaced by $z_{t}^{-}=-z_{t}^{+}, \dot{h}_{0, t}^{-, 1}=-\dot{h}_{0, t}^{+, 1}$ respectively. We can solve these ODEs explicitly. In a first step (regardless of the chosen sign for $z, h_{0}$ )

$$
\hat{Z}_{t}=\left\{\begin{array}{cl}
\sigma_{0} e^{b T}+\frac{a}{b}\left(e^{b t}-1\right) \text { for } b<0 \\
\sigma_{0}+a t & \text { for } b=0
\end{array}\right.
$$

and since

$$
\hat{Y}_{T}^{ \pm}=(p-1) \int_{0}^{T} z_{t}^{ \pm} \hat{Z}_{t} d t
$$

we see that $\hat{Y}_{T}^{-}=-\hat{Y}_{T}^{+}$.In fact, under the (usual) model parameter assumptions $a>0, \sigma_{0}>0$ we see that $\hat{Z}_{t}>0$. We then note that

$$
z_{t}^{ \pm} / q_{0}^{ \pm}=\frac{c^{2}}{\chi_{p}} \sin \left(\chi_{p} t\right) \geq 0 \text { for } t \in[0, T]
$$

indeed we saw that $\chi_{p} T \in[\pi / 2, \pi)$ which implies $\chi_{p} t \in[0, \pi)$ and hence $\sin \left(\chi_{p} t\right) \geq 0$. In particular, given that $q_{0}^{+}>0$ and $p>1$ we see that $\hat{Y}_{T}^{+}>0$ (and then $\hat{Y}_{T}^{-}<0$ ). It follows that

$$
\begin{align*}
c_{2} & :=c_{2}^{+}=\Lambda^{\prime}(1) \times \hat{Y}_{T}^{+, 1} \\
& =p(p-1) \int_{0}^{T} z_{t}^{+} \hat{Y}_{t}^{2} d t \tag{35}
\end{align*}
$$

whereas the contribution from $c_{2}^{-}=\Lambda^{\prime}(1) \times \hat{Y}_{T}^{-, 1}$ is exponentially smaller and will not figure in the expansion. In fact, given the explicit form of $t \mapsto z_{t}^{+}$resp. $\hat{Y}_{t}^{2}$ in terms of $\sin ($.$) and \exp ($.$) ,$ it is clear that the integration in (35) can be carried out in closed form. In doing so, one exploits a cancellation due to

$$
-\chi_{p} \cos \left(\chi_{p} T\right)+b \sin \left(\chi_{p} T\right)=0
$$

and also the equality $\chi_{p}^{2}+b^{2}=c^{2} p(p-1)$, one is led to

$$
c_{2}=q_{0}^{+}\left\{\sigma_{0}+a \frac{\tan \left(\chi_{p} T / 2\right)}{\chi_{p}}\right\} .
$$

It is possible, of course, to substitute the explicitly known quantities $q_{0}^{+}, \chi_{p}$ but this does not yield additional insight.

### 5.2.2 The case of non-zero correlation

We consider again the SDE (20) with diffusion matrix

$$
\sigma=\left(\sigma_{1}, \sigma_{2}\right)=\left(\begin{array}{ll}
z & 0 \\
0 & c
\end{array}\right)
$$

but now allow for correlation $\rho$ between $W^{1}, W^{2}$; we thus have the non-trivial correlation matrix

$$
\Omega=\left(\begin{array}{c}
1 \\
\rho \\
\rho
\end{array}\right) \Longrightarrow \sigma \Omega \sigma^{T}=\left(\begin{array}{cc}
z^{2} & \rho c z \\
\rho c z & c^{2}
\end{array}\right)
$$

In view of financial applications [24] it makes sense to focus on the case $\rho \in(-1,0]$. This will also prove convenient in our analysis below, although there is no doubt that the case $\rho>0$, less interesting in practice, could also be handled within the present framework.

The Hamiltonian becomes, cf. (14),

$$
\begin{aligned}
\mathcal{H}\left(\binom{y}{z} ;(p, q)\right) & =-\frac{1}{2} z^{2} p+b z q+\frac{1}{2}\left(z^{2} p^{2}+c^{2} q^{2}\right)+\rho c z p q \\
& =-\frac{1}{2} z^{2} p+\tilde{b} z q+\frac{1}{2}\left(z^{2} p^{2}+c^{2} q^{2}\right)
\end{aligned}
$$

with

$$
\tilde{b}:=\tilde{b}_{p}:=b+\rho c p
$$

Noting $\partial_{(y, z)} \tilde{b}=(0,0)^{\prime}, \partial_{(p, q)} \tilde{b}=(\rho c, 0)^{\prime}$. The Hamiltonian equations for $\dot{z}, \dot{p}, \dot{q}$, are thus identical as in the uncorrelated case, one just has to replace $b$ by $\tilde{b}$. (In particular, $p_{t}$ is again seen to be constant and we denote its value by $p$.) The Hamiltonian equation for $\dot{y}=\partial_{p} \mathcal{H}$ has, in comparison to the uncorrelated case, an additional term, namely $\left(\partial_{p} \tilde{b}\right) z_{t} q_{t}=\rho c z_{t} q_{t}$. In summary, the Hamiltonian ODEs are

$$
\begin{aligned}
& \binom{\dot{y}_{t}}{\dot{z}_{t}}=\binom{z_{t}^{2}\left(p_{t}-\frac{1}{2}\right)+\rho c z_{t} q_{t}}{\tilde{b} z_{t}+c^{2} q_{t}} \\
& \binom{\dot{p}_{t}}{\dot{q}_{t}}=\binom{0}{p_{t} z_{t}\left(1-p_{t}\right)-\tilde{b} q_{t}} .
\end{aligned}
$$

The following lemma is then obvious (only $y$ requires a computation, due to the additional term in the Hamiltonian ODEs).

Lemma 13 (Partial Hamiltonian Flow, correlated case) Consider the above Hamiltonian ODEs as initial value problem, with initial data $\left(y_{0}, z_{0}\right)=(0,0)$ and $\left(p, q_{0}\right)$ and assume

$$
\begin{equation*}
\chi_{p}^{2}:=c^{2} p(p-1)-\tilde{b}_{p}^{2} \geq 0 \tag{36}
\end{equation*}
$$

Then the explicit solution for $z, p, q$ are then identical to the uncorrelated case, one just has to replace $b$ by $\tilde{b}_{p}$ throughout. The explicit solution for $y$ is modified to

$$
\begin{equation*}
y_{t}=\frac{q_{0}^{2} c^{2}}{8 \chi_{p}^{3}}\left[\left(c^{2}(2 p-1)-2 \rho c \tilde{b}_{p}\right)\left(2 \chi_{p} t-\sin \left(2 \chi_{p} t\right)\right)+2 \rho c \chi_{p}\left(1-\cos \left(2 \chi_{p} t\right)\right)\right] \tag{37}
\end{equation*}
$$

In our explicit analysis of the uncorrelated case (more precisely, in solving the coupled ODEs $\left.\dot{z}_{t}=b z_{t}+c^{2} q_{t}, \dot{q}_{t}=p_{t} z_{t}\left(1-p_{t}\right)-b q_{t}\right)$ we made use of the (model) assumption $b \leq 0$, cf. (33). Conveniently, this remains true when $\rho \in(-1,0]$. Indeed, the following lemma shows we must have $p \geq 0$, so that (with $\rho \leq 0, c>0$ )

$$
\begin{equation*}
\tilde{b}=b+\rho c p \leq 0 \tag{38}
\end{equation*}
$$

Lemma 14 Let $\mathrm{a}>0$. Then $\Lambda(\mathrm{a})=p \mathrm{a}$ and therefore $p \geq 0$.
Proof. We saw in the proof of lemma 7 that, in the uncorrelated case, as a direct consequence of the Hamiltonian ODEs,

$$
p^{2} z_{t}^{2}+c^{2} q_{t}^{2}=2 p \dot{y}_{t}+\partial_{t}\left(z_{t} q_{t}\right)
$$

The correlated case has the identical Hamiltonian ODEs provided we substitute

$$
b \leftarrow \tilde{b} \text { and } \dot{y} \leftarrow \dot{y}-\rho c z_{t} q_{t}
$$

We therefore have

$$
\begin{aligned}
\left|\dot{\mathrm{h}}_{0}(t)\right|^{2} & =\left(\begin{array}{ll}
p & q_{t}
\end{array}\right)\left(\begin{array}{cc}
z^{2} & \rho c z \\
\rho c z & c^{2}
\end{array}\right)\binom{p}{q_{t}}=p^{2} z_{t}^{2}+c^{2} q_{t}^{2}+2 \rho c p z_{t} q_{t} \\
& =2 p\left(\dot{y}_{t}-\rho c z_{t} q_{t}\right)+\partial_{t}\left(z_{t} q_{t}\right)+2 \rho c p z_{t} q_{t}=2 p \dot{y}_{t}+\partial_{t}\left(z_{t} q_{t}\right)
\end{aligned}
$$

and then conclude with the boundary data, exactly as in lemma 7
As already noted, $\tilde{b} \leq 0$ allows to recycle all closed form expressions for $z, q$ obtained in the uncorrelated case - it suffices to replace $b$ by $\tilde{b}$. In particular, for some yet unknown $p, q_{0}$ which may and will depend on $\rho$,

$$
\begin{aligned}
& z_{t}=\frac{q_{0} c^{2}}{\chi_{p}} \sin \left(\chi_{p} t\right) \\
& q_{t}=q_{0}\left(\cos \left(\chi_{p} t\right)-\frac{\tilde{b}}{\chi_{p}} \sin \left(\chi_{p} t\right)\right)
\end{aligned}
$$

where $\chi_{p}^{2}:=c^{2} p(p-1)-\tilde{b}^{2}$ is seen to be positive as in the "uncorrelated" argument. Also, $q_{0} \neq 0$, seen as in the "uncorrelated" case. Transversality, $q_{T}=0$, then implies

$$
\begin{equation*}
\chi_{p} \cos \left(\chi_{p} T\right)-\tilde{b} \sin \left(\chi_{p} T\right)=0 \tag{39}
\end{equation*}
$$

Introducing $r:=\chi_{p} T$ the gives the equation

$$
\begin{equation*}
r \cot r=(b+\rho c p) T \tag{40}
\end{equation*}
$$

On the other hand, from the very definition of $\chi_{p}$, we know

$$
\begin{equation*}
(r / T)^{2}=c^{2} p(p-1)-(b+\rho c p)^{2} . \tag{41}
\end{equation*}
$$

In the uncorrelated case, these two equations were effectively decoupled; in particular, $r \cot r=$ $b T$ lead to $r \in\left\{r_{k}^{+}: k=1,2, \ldots\right\} \subset(0, \infty)$, written in increasing order. Since $p^{+}$was seen to be monotonically increasing in $r$, cf. equation (29), and we were looking for the minimal $p$, corresponding to the minimal energy (cf. lemma 14), we were led to seek the first positive root $r_{1}^{+}$. (In fact, $r_{1}^{+} \in(\pi / 2, \pi)$ as we will also find in the "correlated" discussion below.)

The correlated case is a little more complicated and we start in expressing $p$ in equation (40) in terms of $r$. Indeed, the quadratic equation (41) shows

$$
\begin{equation*}
p^{ \pm}(r)=\frac{1}{2\left(1-\rho^{2}\right)}\left\{\left(1+2 \rho \frac{b}{c}\right) \pm \sqrt{\left(1+2 \rho \frac{b}{c}\right)^{2}+4\left(1-\rho^{2}\right)\left[\frac{b^{2}}{c^{2}}+\frac{r^{2}}{c^{2} T^{2}}\right]}\right\} \tag{42}
\end{equation*}
$$

where $p^{-}(r)<0$ (and hence can be ignored in view of lemma 14) and $p^{+}(r)>0$. We now look
for $r$ which satisfies the equation

$$
r \cot r=\left(b+\rho c p^{+}(r)\right) T
$$

It is elementary to see that $r \cot r$ is non-negative on $[0, \pi / 2]$ and then maps $[\pi / 2, \pi)$ strictly monotonically to $(-\infty, 0]$. On the other hand, the map $r \mapsto\left(b+\rho c p^{+}(r)\right) T$ is $\leq 0$ for all $r$;
in particular, there will be a first intersection with the graph of $r \mapsto r \cot r$ in $[\pi / 2, \pi)$, say at $r=r_{1}^{+}$. Since $p^{+}(r)$ is plainly strictly increasing in $r$, the minimal $p$ must equal to

$$
p_{1}^{+}:=p^{+}\left(r_{1}^{+}\right) .
$$

We then proceed as in the uncorrelated case, and determine $q_{0}$ from the boundary condition $y_{T}=\mathrm{a}>0$ where $y$ is now given by (37). This leads to $q_{0} \in\left\{q_{0,1}^{+}, q_{0,1}^{-}\right\}$where

$$
q_{0,1}^{ \pm}= \pm \frac{2}{c} \sqrt{\frac{2 r^{3} \mathrm{a}}{T^{3}\left(\left(c^{2}(2 p-1)-2 \rho c \tilde{b}\right)(2 r-\sin (2 r))+2 \rho c r / T(1-\cos (2 r))\right)}}
$$

where $r=r_{1}^{+}$and $p=p_{1}^{+}$. Again, we have two minimizing controls, $\mathcal{K}_{\mathrm{a}}^{\text {min }}=\left\{\mathrm{h}_{0}^{+}, \mathrm{h}_{0}^{-}\right\}$. We now have

$$
\dot{\mathrm{h}}_{0}(t)=\left(\begin{array}{cc}
z_{t} \sqrt{1-\rho^{2}} & 0  \tag{43}\\
\rho z_{t} & c
\end{array}\right)\binom{p}{q_{t}}
$$

instead of (26) and of course lemma 13 implies that $z_{t}$ and $q_{t}$ are fully and explicitly determined for each choice of $\left(p, q_{0}\right)$. In particular for $\left(p, q_{0}\right) \leftarrow\left(p_{1}^{+}, q_{0,1}^{+}\right)$resp. $\left(p_{1}^{+}, q_{0,1}^{-}\right)$we so obtain $\mathrm{h}_{0}^{+}$ resp. $\mathrm{h}_{0}^{-}$which can be written explicitly by simple substitution. Moreover, and again as in the uncorrelated case,

$$
\begin{equation*}
\Lambda(\mathrm{a})=\frac{1}{2}\left\|\mathrm{~h}_{0}^{+}\right\|_{H}^{2}=\frac{1}{2}\left\|\mathrm{~h}_{0}^{-}\right\|_{H}^{2}=p_{1}^{+} \mathrm{a} \tag{44}
\end{equation*}
$$

and upon taking $\mathrm{a}=1$ we have computed the leading order constant

$$
c_{1}=\Lambda(1)=p_{1}^{+} \equiv p^{+}\left(r_{1}^{+}\right)
$$

where we recall that $r_{1}^{+}$is the first intersection point of $r \mapsto r \cot r$ with $\left(b+\rho c p^{+}(r)\right) T$ and $p^{+}(\cdot)$ was given in (42).

At last, we turn to the computation of the second-order exponential constant, $c_{2}$. As in the uncorrelated case, we ease notation by writing $\left(p, q_{0}^{ \pm}\right)$instead of $\left(p_{1}^{+}, q_{0,1}^{+}\right)$resp. $\left(p_{1}^{+}, q_{0,1}^{-}\right)$ for the rest of this section. Again, we have to consider ODEs for $\left(\hat{Y}_{t}, \hat{Z}_{t}\right)$, for each minimizer, $\mathrm{h}_{0}^{+}=\left(h_{0, \cdot}^{+, 1}, h_{0,{ }^{+}}^{+, 2}\right)$ and $\mathrm{h}_{0}^{-}=\left(h_{0,{ }^{+},}^{+, 1},-h_{0, \cdot}^{+, 2}\right)$. Recall from (43) that, with $\bar{\rho}=\sqrt{1-\rho^{2}}$,

$$
\dot{\mathrm{h}}_{0}^{+}(t)=\binom{p \bar{\rho} z_{t}^{+}}{\rho p z_{t}^{+}+c q_{t}^{+}}
$$

where $(\cdot)^{ \pm}$indicates the chosen sign of $q_{0} \in\left\{q_{0,1}^{+}, q_{0,1}^{-}\right\}$which determines the choice of minimizer. We first determine $\hat{Y}_{T}=\hat{Y}_{T}\left(\mathrm{~h}_{0}^{+}\right)$from the ODE

$$
\begin{aligned}
\frac{d}{d t}\binom{\hat{Y}_{t}}{\hat{Z}_{t}^{2}} & =\left\{\left(\begin{array}{cc}
0 & -z_{t}^{+} \\
0 & b
\end{array}\right)+\left(\begin{array}{ll}
0 & \bar{\rho} \\
0 & 0
\end{array}\right) \dot{h}_{0, t}^{+, 1}+\left(\begin{array}{ll}
0 & \rho \\
0 & 0
\end{array}\right) \dot{h}_{0, t}^{+, 2}\right\}\binom{\hat{Y}_{t}}{\hat{Z}_{t}}+\binom{0}{a} \\
& =\left(\begin{array}{cc}
0 & (p-1) z_{t}^{+}+\rho c q_{t}^{+} \\
0 & b
\end{array}\right)\binom{\hat{Y}_{t}}{\hat{Z}_{t}}+\binom{0}{a} \\
\text { with }\binom{\hat{Y}_{0}}{\hat{Z}_{0}^{2}} & =\binom{0}{\sigma_{0}} .
\end{aligned}
$$

This already shows that we have the identical (closed form) ODE solution for $\hat{Z}_{t}$ as in the uncorrelated case. On the other hand, the form of $\hat{Y}_{T}$ now exhibits an additional term as is seen in

$$
\hat{Y}_{T}=(p-1) \int_{0}^{T} z_{t}^{+} \hat{Z}_{t} d t+\rho c \int_{0}^{T} q_{t}^{+} \hat{Z}_{t} d t
$$

Since $q_{t}^{+}$is essentially of the same trigonometric form as $z_{t}^{+}$, it is clear that the explicit computations of the uncorrelated case extend. In the end, one finds without too much difficulties

$$
c_{2}^{+}=\Lambda^{\prime}(1) \times \hat{Y}_{T}\left(\mathrm{~h}_{0}^{+}\right)=q_{0}^{+}\left\{\sigma_{0}+a \frac{\tan \left(\chi_{p} T / 2\right)}{\chi_{p}}\right\} .
$$

A similar computation along $\mathrm{h}_{0}^{-}$gives $c_{2}^{+}=\Lambda^{\prime}(1) \times \hat{Y}_{T}\left(\mathrm{~h}_{0}^{-}\right)$in explicit form and $c_{2}$ is identified as max $\left(c_{2}^{+}, c_{2}^{-}\right)$.

### 5.2.3 Checking non-degeneracy, zero and non-zero correlation

We now check the non-degeneracy conditions, contained in assumptions (i)-(iii) of theorem 1 , which of course is the ultimate justification that an expansion of the form (21) with the constants computed above holds true. Again, focus is on the case of correlation parameter $\rho \in(-1,0]$. We saw in the previous sections (for $\rho=0$, then $\rho \leq 0$ ) that $\# K_{\mathrm{a}}^{\min }=\#\left\{\mathrm{~h}_{0}^{+}, \mathrm{h}_{0}^{-}\right\}=2$, whenever $\mathrm{a}>0$. (In fact, we apply this with $\mathrm{a}=1$.)

Secondly, a look at (23) reveals that the degenerate region is $\{(y, z): z=0\}$, the complement of which is elliptic. Clearly, no controlled path which reaches $y_{T}=\mathrm{a}>0$ can stay in the degenerate region for all times $t \in[0, T]$; after all, this would entail $d y=0$ and hence $y_{T}=0$. We conclude the any ODE solution driven by $h \in \mathcal{K}_{a}$ must intersect the region of ellipticity; but this already implies non-degeneracy of the corresponding (deterministic) Malliavin covariance matrix.

At last, we check non-focality and focus on $\mathrm{h}_{0}^{+}$, the other case being similar. We have to check non-degeneracy of the Jacobian of the map $\pi H_{0 \leftarrow T}(\mathrm{a}, \cdot ; *, 0)$, evaluated at $\cdot=z_{T}, *=p_{T}$
after differentiation, where $z_{T}, p_{T}$ are obtained form the Hamiltonian flow at time $T$, cf. lemma [13, with time 0 initial data $\left(0,0 ; p_{1}^{+}, q_{0,1}^{+}\right)$.

With some abuse of notation, write

$$
\binom{y_{0}}{z_{0}} \equiv\binom{y_{0}(z, p)}{z_{0}(z, p)} \equiv \pi H_{0 \leftarrow T}(\mathrm{a}, z ; p, 0) .
$$

Our non-degeneracy condition requires us to show that

$$
\left.\operatorname{det}\left(\begin{array}{ll}
\frac{\partial y_{0}}{\partial p} & \frac{\partial y_{0}}{\partial z}  \tag{45}\\
\frac{\partial z_{0}}{\partial p} & \frac{\partial z_{0}}{\partial z}
\end{array}\right)\right|_{*} \neq 0
$$

where $\left.(\ldots)\right|_{*}$ indicates evaluation $\left.(\ldots)\right|_{(p, z)=\left(p^{+}, z_{T}\right)}$ in the sequel. This implies in particular that all expressions which are formulated in terms of the solutions to the Hamiltonian flows, reduced to the corresponding expressions identified in proposition 11 for $\rho=0$, resp. in section 5.2.2 for $\rho \leq 0$. For instance, $\left.\left(y_{0}, z_{0}\right)\right|_{*}=(0,0),\left.y_{T}\right|_{*}=\mathrm{a},\left.z\right|_{*}=z_{T} \neq 0,\left.\chi_{p} T\right|_{*} \in[\pi / 2, \pi)$ and so.

Since (z., q.) solves a linear ODE, we can compute

$$
\begin{aligned}
z_{0}(z, p) & =(10) e^{-T\left(\begin{array}{cc}
\tilde{b}_{p} & c^{2} \\
p(1-p)-\tilde{b}_{p}
\end{array}\right)\binom{z}{0}} \\
& =\frac{z}{\chi_{p}}\left(\chi_{p} \cos \left(\chi_{p} T\right)-\tilde{b}_{p} \sin \left(\chi_{p} T\right)\right) .
\end{aligned}
$$

We first note that $\partial z_{0} /\left.\partial z\right|_{*}$ is zero; indeed, this follows from (39). Our next claim is $\partial y_{0} /\left.\partial z\right|_{*} \neq 0$. Indeed, from the structure of the Hamilton ODEs,

$$
y_{0}-\mathrm{a}=-\int_{0}^{T} \dot{y}_{t} d t=z^{2}(\ldots)
$$

where $(\cdots)$ does not depend on $z$. As a result $\partial y_{0} /\left.\partial z\right|_{*}=\left.2 z(\ldots)\right|_{*}=\left.2 \frac{y_{0}-\mathrm{a}}{z}\right|_{*}=-2 \mathrm{a} / z_{T} \neq 0$.
It remains to check that $\partial z_{0} /\left.\partial p\right|_{*} \neq 0$. To this end, recall, as a consequence of the transversality condition, see (39), that $\chi_{p} \cos \left(\chi_{p} T\right)-\left.\tilde{b}_{p} \sin \left(\chi_{p} T\right)\right|_{*}=0$. It follows that

$$
\partial z_{0} /\left.\partial p\right|_{*}=\left\{\frac{z}{\chi_{p}} \frac{\partial}{\partial p}\left(\chi_{p} \cos \left(\chi_{p} T\right)-\tilde{b}_{p} \sin \left(\chi_{p} T\right)\right)\right\}_{*}
$$

and since $z /\left.\chi_{p}\right|_{*} \neq 0$, it will be enough to show (strict) negativity of $\left.\frac{\partial}{\partial p}(\ldots)\right|_{*}$ above. By scaling, there is no loss of generality in taking $T=1$ and we shall do so from here on. Then

$$
\begin{aligned}
& \frac{\partial}{\partial p}\left(\chi_{p} \cos \left(\chi_{p}\right)-\tilde{b}_{p} \sin \left(\chi_{p}\right)\right) \\
= & \chi_{p}^{\prime}\left[\left(1-\tilde{b}_{p}\right) \cos \left(\chi_{p}\right)-\chi_{p} \sin \left(\chi_{p}\right)\right]-\rho c \sin \left(\chi_{p}\right) .
\end{aligned}
$$

Since $\left.\tilde{b}_{p}\right|_{*} \leq 0$ and $\left.\chi_{p}\right|_{*} \in[\pi / 2, \pi)$ we see that $\left.[\ldots]\right|_{*}<0$. Given that $\left.\chi_{p}^{\prime}\right|_{*}>0$, this already settles the negativity claim in the zero-correlation case. In the case $-1<\rho<0$, we use (39) to write

$$
\begin{aligned}
& \left.\frac{\partial}{\partial p}\left(\chi_{p} \cos \left(\chi_{p}\right)-\tilde{b}_{p} \sin \left(\chi_{p}\right)\right)\right|_{*} \\
= & \chi_{p}^{\prime}\left[\left(1-\tilde{b}_{p}\right) \frac{\tilde{b}_{p} \sin \left(\chi_{p}\right)}{\chi_{p}}-\chi_{p} \sin \left(\chi_{p}\right)\right]-\left.\rho c \sin \left(\chi_{p}\right)\right|_{*} .
\end{aligned}
$$

After division by $\sin \left(\chi_{p}\right) /\left.\chi_{p}\right|_{*}>0$, we have, using $\tilde{b}_{p}=b+\rho c p \leq 0, b \leq 0$ and again $\left.\chi_{p}^{\prime}\right|_{*}>0$,

$$
\begin{aligned}
& \chi_{p}^{\prime}\left[\left(1-\tilde{b}_{p}\right) \tilde{b}_{p}-\chi_{p}^{2}\right]-\left.\rho c \chi_{p}\right|_{*} \\
\leq & \chi_{p}^{\prime}\left[(1-\rho c p) \rho c p-\chi_{p}^{2}\right]-\left.\rho c \chi_{p}\right|_{*} \\
\leq & -\left.\rho c\left(\chi_{p}-p \chi_{p}^{\prime}\right)\right|_{*} .
\end{aligned}
$$

With $-\rho c>0$, it will then be sufficient to show strict negativity of $\chi_{p}-\left.p \chi_{p}^{\prime}\right|_{*}$. To this end note that the definition, $\chi_{p}^{2}=c^{2} p(p-1)-\tilde{b}^{2}$, implies

$$
\begin{aligned}
2 \chi_{p} \chi_{p}^{\prime} & =c^{2}(2 p-1)-2 \tilde{b}(\rho c) \\
\chi_{p} p \chi_{p}^{\prime} & =c^{2} p(p-1 / 2)-\tilde{b}(\rho c p) \\
& =\chi_{p}^{2}+\frac{c^{2} p}{2}+b \tilde{b}>\chi_{p}^{2}
\end{aligned}
$$

whenever $c^{2} p / 2+b \tilde{b}>0$ which is surely the case upon evaluation $\left.\ldots\right|_{*}$.
We conclude that $\partial z_{0} /\left.\partial p\right|_{*} \neq 0$, and then validity of (45), for any parameter set $\rho \in$ $(-1,0], b \leq 0, c>0, T>0$. In other words, we have completed the check of our non-degeneracy condition.

### 5.3 Comments on Heston [33] and Lions-Musiela [39]

We recall from [28, 21] that the density of log-stock price $Y_{T}$ in the Heston model,

$$
\begin{aligned}
d Y & =-V / 2+\sqrt{V} d W^{1}, X(0)=x_{0}=0 \\
d V & =(a+b V) d t+c \sqrt{V} d W^{2}, \quad V(0)=v_{0}>0
\end{aligned}
$$

with $a \geq 0, b \leq 0, c>0$ and correlation $\rho \in(-1,0]$ has the form

$$
f(y)=e^{-c_{1} y} e^{c_{2} \sqrt{y}} y^{-3 / 4+a / c^{2}}\left(c_{3}+O(1 / \sqrt{y})\right) \text { as } y \rightarrow \infty
$$

with explicitly computable $c_{1}=C_{1}(b, c, \rho, T)$ and $c_{2}=\sqrt{v_{0}} \times C_{2}(b, c, \rho, T)$, both do not depend on $a$. While scaling with $\theta=2$,

$$
Y_{\varepsilon}:=\varepsilon^{2} Y, \quad V_{\varepsilon}:=\varepsilon^{2} V
$$

indeed yields a small noise problem, namely

$$
\begin{aligned}
d Y^{\varepsilon} & =-V^{\varepsilon} / 2+\sqrt{V^{\varepsilon}} \varepsilon d W^{1}, \quad X(0)=x_{0}=0 \\
d V^{\varepsilon} & =\left(a \varepsilon^{2}+b V^{\varepsilon}\right) d t+c \sqrt{V^{\varepsilon}} \varepsilon d W^{2}, \quad V(0)=v_{0} \varepsilon^{2}>0
\end{aligned}
$$

The algebraic factor $y^{-3 / 4+a / c^{2}}$ in the above expansion then contradicts the expected factor; cf. (13)

$$
y^{\frac{1}{\theta}-1}=y^{-1 / 2}
$$

There is no contradiction here, of course. Rather, we see an explicit example where "formal" application of a theorem to a model which is short of the required regularity leads to wrong conclusion (at least at the fine level of algebraic factors). Remark that one can trace the origin of this unexpected $y^{-3 / 4+a / c^{2}}$ factor to the behaviour of the one-dimensional variance process $V$; also known as Feller - or Cox-Ingersoll-Ross diffusion. Curiously then even a large deviation
principle for $V^{\varepsilon}$ as given above presently lacks justification, despite the recent advances in 19, 6]. Clearly then, we are not anywhere near in obtaining the Heston tail result of 28, 21 with the present methods.

However, in the special case when $a=c^{2} / 4$ it is an easy exercise to see that the Heston model can be realized as Stein-Stein model (take $V=Z^{2}$, where $Z$ is the volatility component of the Stein-Stein model), the resulting expressions are then seen to be consistent with those obtained in 21] and, in particular, $y^{-3 / 4+a / c^{2}}=y^{-1 / 2}$.

Another class of non-smooth, non-affine stochastic vol model with " $\theta=2$ "-scaling, introduced by Lions-Musiela [39]. For $\delta \in[1 / 2,1]$ and $\gamma=1-\delta$ they consider the 2-dimensional diffusion

$$
\begin{aligned}
d Y & =-\frac{1}{2} Z^{2 \delta} d t+Z^{\delta} d \tilde{W}_{1}, \quad Y_{0}
\end{aligned}=0 .
$$

And indeed with $Y_{\varepsilon}=\varepsilon^{2} Y$ and $Z_{\varepsilon}=\varepsilon^{1 / \delta} Z$ this becomes a small noise problem;

$$
\begin{aligned}
& d Y_{\varepsilon}=-\frac{1}{2} Z_{\varepsilon}^{2 \delta} d t+Z_{\varepsilon}^{\delta} \varepsilon d W, \quad Y_{\varepsilon}(0)=0 \\
& d Z_{\varepsilon}=b Z_{\varepsilon} d t+c Z_{\varepsilon}^{\gamma} \varepsilon d Z, \quad Z_{\varepsilon}(0)=\varepsilon^{1 / \delta} z_{0}
\end{aligned}
$$

In their paper they establish exponential moments of $Y_{T}$. It is tempting to use corollary 5. at least to leading large deviation order, to obtain the exponential tail of $Z$ for models that scale with $\theta=2$. Of course, as was discussed in the Heston case, such a "formal" application can be wrong. Further work, building on [19], 6], will be necessary to deal with such degenerate models directly.

## References

[1] Y. Ait-Sahalia, Closed-form likelihood expansions for multivariate diffusions. The Annals of Statistics 2008, Vol. 36, No. 2, 906-937
[2] M. Avellaneda, D. Boyer-Olson, J.Busca, P. Friz: Application of large deviation methods to the pricing of index options in finance, Comptes Rendus de l'Académie des Sciences - Series I-Mathematique (2003).
[3] M. Avellaneda, D. Boyer-Olson, J.Busca, P. Friz: Reconstructing volatility, RISK (2004).
[4] R. Azencott. Formule de Taylor stochastique et développement asymptotique d'intégrales de Feynmann. Séminaire de Probabilités XVI; Supplément: Géométrie différentielle stochastique. Lecture notes in Mathematics, 921, 237-285, 1982.
[5] R. Azencott. Petites perturbations aléatoires des systèmes dynamiques: développements asymptotiques. Bulletin des sciences mathématiques. vol. 109, no3, pp. 253-308, 1985.
[6] P. Baldi and L. Caramellino. General Freidlin-Wentzell large deviations and positive diffusions. Forthcoming in Statistics and Probability Letters, 2011.
[7] Benaim, Friz: Regular Variation and Smile Asymptotics, Math. Finance Vol. 19 no 1. (2009), 1-12
[8] S. Benaim and P. Friz. Smile asymptotics II: models with known moment generating functions. Journal of Applied Probability, 45: 16-32, 2008.
[9] G. Ben Arous. Développement asymptotique du noyau de la chaleur hypoelliptique hors du cut-locus. Annales Scientifiques de l'Ecole Normale Supérieure, 4 (21): 307-331, 1988.
[10] G. Ben Arous. Methods de Laplace et de la phase stationnaire sur l'espace de Wiener. Stochastics, 25: 125-153, 1988.
[11] G. Ben Arous, P. Laurence: Second order expansion for implied volatility in two factor local-stochastic volatility models and applications to the dynamic Sabr model. Preprint 2010.
[12] G. Ben Arous, R. Léandre: Décroissance exponentielle du noyau de la chaleur sur la diagonale (I), Probab. Th. Re1. Fields 90, 175-202 (1991)
[13] R. Bishop, R. Crittenden, Geometry of Manifolds, Academic Press 1964.
[14] H. Berestycki, J. Busca, and I. Florent. Computing the implied volatility in stochastic volatility models. Communications on Pure and Applied Mathematics, 57(10):1352-1373, 2004.
[15] J.M. Bismut. Malliavin Calculus and Large Deviations. 1984
[16] P Bourgade and O Croissant. Heat kernel expansion for a family of stochastic volatility models : -geometry; arXiv:cs.CE/0511024 2005.
[17] J.D. Deuschel, P.K. Friz, A. Jacquier, S. Violante. Marginal density expansions for diffusions and stochastic volatility, part I: Theoretical Foundations. Communications on Pure and Applied Mathematics, to appear.
[18] J.D. Deuschel and D.W. Stroock. Large Deviations. Volume 342 of AMS/Chelsea Series. 2000
[19] C. Donati-Martin, A. Rouault, M. Yor and M. Zani. Large deviations for squares of Bessel and Ornstein-Uhlenbeck processes. PTRF, 129(2), 261-289.
[20] M. Freidlin and A.D. Wentzell. Random perturbations of dynamical systems. Grundlehren der Mathematischen Wissenschaften (Second edition ed.). New York: Springer-Verlag, 1998.
[21] P. Friz, S. Gerhold, A. Gulisashvili and S. Sturm. Refined implied volatility expansions in the Heston model. Quant. Finance, Volume 11, Issue 8, 1151-1164, 2011.
[22] Friz, Peter; Victoir, Nicolas; Multidimensional Stochastic Processes as Rough Paths. Theory and Applications, Cambridge Studies of Advanced Mathematics Vol. 120, 670 p., Cambridge University Press
[23] K. Gao and R. Lee. Asymptotics of implied volatility to arbitrary order. Preprint available at http://ssrn.com/abstract=1768383, 2011.
[24] Gatheral, Jim; The Volatilty Surface. Wiley Finance, 2006.
[25] Gatheral, Jim; Further Developments in Volatility Derivatives Modeling. Presentation 2008. Available on www.math.nyu.edu/fellows_fin.../gatheral/FurtherVolDerivatives2008.pdf
[26] Gatheral, Jim; Hsu, Elton P.; Laurence, Peter; Ouyang, Cheng; Wang, Tai-Ho. Asymptotics of Implied Vol in Local Vol Models. Math. Finance, Volume 22, Issue 4, pages 591-620, October 2012.
[27] J. Gatheral and A. Jacquier, Convergence of Heston to SVI, Quant. Finance 2011, Volume 11, Issue 8, 2011.
[28] A. Gulisashvili and E. Stein. Asymptotic Behavior of the Stock Price Distribution Density and Implied Volatility in Stochastic Volatility Models, Applied Mathematics \& Optimization, Volume 61, Number 3, 287-315, DOI: 10.1007/s00245-009-9085-x
[29] A. Gulisashvili. Analytically tractable stochastic volatility models, Springer Finance, 2012.
[30] I. Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. Probab. Theory Relat. Fields, 71(4):501-516, 1986
[31] Patrick Hagan, Andrew Lesniewski, and Diana Woodward; Probability Distribution in the SABR Model of Stochastic Volatility. Working Paper 2005. Available on lesniewski.us/working.html
[32] Henry-Labordère P, Analysis, geometry and modeling in finance, Chapman and Hill/CRC, 2008.
[33] Heston S. 1993. A closed-form solution for options with stochastic volatility, with application to bond and currency options. Review of Financial Studies 6, 327-343.
[34] Jurdjevic, Kupka; Polynomial Control Systems; Math. Ann. 272, 361-368 (1985)
[35] Yu. I. Kifer, "On the asymptotics of the transition probability density of processes with small diffusion", Teor. Veroyatnost. i Primenen., 21:3 (1976), 527-536
[36] Shigeo Kusuoka and Yasufumi Osajima: A remark on the asymptotic expansion of density function of Wiener functionals. UTMS Preprint 2007-18.
[37] Shigeo Kusuoka and Daniel W. Stroock, Precise asymptotics of certain Wiener functionals. Journal of Functional Analysis, Volume 99, Issue 1, July 1991, Pages 1-74.
[38] Roger Lee, "The Moment Formula for Implied Volatility at Extreme Strikes " Mathematical Finance, vol 14 issue 3 (July 2004), 469-480.
[39] P.L. Lions and M. Musiela. Correlations and bounds for stochastic volatility models. Ann. I.H. Poincaré, 24, 2007, 1-16.
[40] Alex Lipton and Artur Sepp, Stochastic volatility models and Kelvin waves. 2008, J. Phys. A: Math. Theor. 41.
[41] Alex Lipton, Presentation at Global Derivatives, Barcelon 2012.
[42] S A Molchanov, "Diffusion processes and Riemannian geometry", Russ. Math. Surv., 1975, 30 (1), 1-63.
[43] R. Montgomery. A Tour of SubRiemannian Geometries, their Geodesics and Applications, Volume 91 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2002.
[44] Osajima, Yasufumi, General Asymptotics of Wiener Functionals and Application to Mathematical Finance (July 25, 2007). Available at SSRN: http://ssrn.com/abstract=1019587
[45] Paulot, Louis, Asymptotic Implied Volatility at the Second Order with Application to the SABR Model (June 3, 2009). Available at SSRN: http://ssrn.com/abstract=1413649
[46] Huyen Pham, Large deviations in Finance, 2010, Third SMAI European Summer School in Financial Mathematics.
[47] V. Piterbarg, Markovian projection method for volatility calibration; RISK (2007).
[48] Sakai, T.: Riemannian Geometry, AMS, 1992.
[49] R. Schöbel and J. Zhu, Stochastic volatility with an OrnsteinUhlenbeck process: An Extension, European Finance Review, 3 (1999), pp. 2346.
[50] Seierstad, A. and Sydsaeter, K.: Optimal Control Theory with Economic Applications. (Advanced Textbooks in Economics, 24). North- Holland Amsterdam, 1987
[51] Stein, E. M., and J. C. Stein, 1991, "Stock Price Distributions with Stochastic Volatility: An Analytic Approach," Review of Financial Studies, 4, 727-752.
[52] Takanobu S. Watanabe S.: Asymptotic expansion formulas of the Schilder type for a class of conditional Wiener functional integration. In "Asymptotics problems in probability theory: Wiener functionals and asymptotics". K.D. Elworthy N. Ikeda edit. Pitman. Res. Notes. Math. Series. 284 (1993), 194-241.
[53] Varadhan, S. R. S., On the behavior of the fundamental solution of the heat equation with variable coefficients. Communications on Pure and Applied Mathematics, 20: 431-455. 1967
[54] Varadhan, S.R.S.: Lectures on large deviations, available at http://math.nyu.edu/faculty/varadhan/LDP.html, 2010.


[^0]:    ${ }^{1}$ Sometimes the Stein-Stein model is written with $|Z| d W^{1}$ rather than $Z d W^{1}$; in the uncorrelated case this does not make a difference to the law of the process, as is immediate from a look at the respective generators. There is a recent tendency in the finance community to use the form $Z d W^{1}$ which we analyze here, cf. 40, 41, this version of the model was also proposed by Schöbel-Zhu, 49.
    ${ }^{2}$ Strictly speaking, the $O$-term given in [28] is $\log s$ with power $-1 / 4$; the authors have informed us, however, that a closer look at their argument indeed gives power $-1 / 2$.
    ${ }^{3}$ Small strike asymptotics are similar and will not be discussed here.

[^1]:    ${ }^{4}$ In contrast, short time asymptotics in "locally elliptic" stochastic volatility models usually can be localized to a non-degenerate region.

[^2]:    ${ }^{5}$ If (6) is understood in Stratonovich sense, so that $d W$ is replaced by $\circ d W$, the drift vector field $b(\varepsilon, \cdot)$ is changed to $\tilde{b}(\varepsilon, \cdot)=b(\varepsilon, \cdot)-\left(\varepsilon^{2} / 2\right) \sum_{i=1}^{m} \sigma_{i} \cdot \partial \sigma_{i}$. In particular, $\sigma_{0}$ is also the limit of $\tilde{b}(\varepsilon, \cdot)$ in the sense of (7).

[^3]:    ${ }^{6}$ A well-known sufficient condition (cf. [17] and the references therein) is the strong Hörmander condition (H1), as stated in corollary 4 below.

[^4]:    ${ }^{7}$ We have $\mathrm{a}=\mathrm{y} \in \mathbb{R}^{l}$, in context of small noise and short time expansions, and $\mathrm{a}=1 \in \mathbb{R}^{l}$, with $l=1$, in the context of tail expansions, corollary 5

[^5]:    ${ }^{8}$ All explicit solutions given in (28) are even functions of $\chi_{p_{0}}$ and have a removable singularity for $\chi_{p_{0}}=0$. By convention we shall always assume $\chi_{p_{0}} \geq 0$ although the sign of $\chi_{p_{0}}$ does not matter.

