CENTER FOR ADVANCED RESEARCH IN FINANCE

C A R F W o r k i n g P a p e r

CARF-F-289

Momentum-Space Approach to Asymptotic Expansion for Stochastic Filtering and other Problems

Masaaki Fujii
The University of Tokyo

September 2012


CARF is presently supported by Bank of Tokyo-Mitsubishi UFJ, Ltd., Dai-ichi Mutual Life Insurance Company, Meiji Yasuda Life Insurance Company, Nomura Holdings, Inc. and Sumitomo Mitsui Banking Corporation (in alphabetical order). This financial support enables us to issue CARF Working Papers.

CARF Working Papers can be downloaded without charge from:
http://www.carf.e.u-tokyo.ac.jp/workingpaper/index.cgi

Momentum-Space Approach to Asymptotic Expansion for Stochastic Filtering and other Problems *

Masaaki Fujii†

First version: September 10, 2012

Abstract

This paper develops an asymptotic expansion technique in momentum space. It is shown that Fourier transformation combined with a polynomial-function approximation of the nonlinear terms gives a closed recursive system of ordinary differential equations (ODEs) as an asymptotic expansion of the conditional distribution appearing in stochastic filtering problems. Thanks to the simplicity of the ODE system, higher order calculation can be performed easily. Furthermore, solving ODEs sequentially with small sub-periods with updated initial conditions makes it possible to implement a *substepping method* for asymptotic expansion in a numerically efficient way. This is found to improve the performance significantly where otherwise the approximation fails badly. The method may be useful for other applications, such as, option pricing in finance as well as measure-valued stochastic dynamics in general.

Keywords : Asymptotic Expansion, Fourier Transformation, Filtering, Zakai equation, polynomial-function approximation, measure-valued nonlinear system

1 Introduction

In many areas, researchers frequently encounter the situation where crucial parameters for their models are not directly observable in our mother nature or in experiments. This is particularly the case, for examples, in engineering, applied physics, finance and economics. To get the best estimate of the unobservable from what we can directly observe is the goal of stochastic filtering. The most famous example with analytical solution is Kalman-Bucy filter, which assumes both of the signal and observation processes are linear and hence associated with Gaussian distributions.

However, there are many cases where interested variables follow nonlinear stochastic processes and their distributions are far from Gaussian. This is particularly the case

*This research is supported by CARF (Center for Advanced Research in Finance) and the global COE program "The research and training center for new development in mathematics." All the contents expressed in this research are solely those of the authors and do not represent any views or opinions of any institutions. The authors are not responsible or liable in any manner for any losses and/or damages caused by the use of any contents in this research.

†Graduate School of Economics, The University of Tokyo

for financial problems. In fact, many people were forced to realize the sheer impacts of non-Gaussianity in the last financial crisis followed by the collapse of Lehman Brothers. Researchers, practitioners as well as regulators now clearly recognize the importance of understanding not only the first two moments but also every other details of relevant distributions. Here, we need to deal with nonlinear filtering problems. Filtering theory has a long history and is still developing very rapidly, partly helped by the great increase of computational power. Recently, there appeared a thick volume [2] from Oxford university press, which contains latest developments and reviews for theoretical as well as numerical techniques for nonlinear filtering problems. Unfortunately though, they seem to require very sophisticated mathematical as well as numerical skills and there exist many hurdles to cross for non-specialists.

In this paper, we propose a simple approximation scheme based on an asymptotic expansion method in momentum space for nonlinear filtering problems. The method should be also useful for other financial problems that do not require filtering. Widely used "position-space" asymptotic expansion method (See [5], and references therein.) is transformed into a simpler form in the momentum (or Fourier transformed) apace, and the resultant dynamics of the characteristic function is given by a closed recursive system of ordinary differential equations (ODEs). It is shown that the form of ODEs unchanged for any order of expansion, which allows straightforward numerical implementation for higher order approximations. Furthermore, dividing the original time horizon into a set of small sub-periods and solving the ODEs sequentially with updated initial conditions, which we call *substepping method* for asymptotic expansion, increases the parameter space where the approximation is effective. Two simple examples are discussed to demonstrate how the method works. We also make a brief comment on the possibility that the same method can be used to analyze other measure-valued nonlinear systems.

2 Preliminaries for Nonlinear Filtering

2.1 Zakai equation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions. We consider n-dimensional signal process $X = \{X_t, t \geq 0\}$ and m-dimensional observation process $Y = \{Y_t, t \geq 0\}$ following the dynamics of

$$X_t = \mu(t, X_t)dt + \eta(t, X_t)dV_t + \bar{\eta}(t, X_t)dW_t \tag{2.1}$$

$$Y_t = h(t, X_t)dt + dW_t \tag{2.2}$$

with $Y_0 = 0$ and an independent initial distribution for X_0. Here V and W are independent $(\mathbb{P}, \mathcal{F})$-Brownian motions with dimensionality d and m, respectively. μ, h, η and $\bar{\eta}$ are deterministic function of (t, x) [1] and take values in $\mathbb{R}^n, \mathbb{R}^m, \mathbb{R}^{n\times d}$ and $\mathbb{R}^{n\times m}$, respectively. The functions μ, η and $\bar{\eta}$ are assumed to satisfy appropriate conditions so that (2.1) has a unique solution. The measurable function h is assumed to satisfy the conditions that makes the following process $Z = \{Z_t, t \geq 0\}$ be a martingale:

$$Z_t = \exp\left(-\int_0^t h_s(X_s)^\top dW_s - \frac{1}{2}\int_0^t ||h_s(X_s)||^2 ds\right) \tag{2.3}$$

[1] For simplicity, we frequently use a notation of $\mu_t(X_t)$ and similarly for other functions.

where $\top$ denotes the transposition. We denote $\{\mathcal{Y}_t, t \geq 0\}$ be the usual augmented filtration generated by the process Y. Our goal of the filtering problem is to obtain the conditional distribution π_t of the signal X at time t given the information available from observing the process Y in the interval of $[0, t]$. In other words, for a given arbitrary bounded function φ, computing

$$\pi_t(\varphi) = \mathbb{E}\Big[\varphi(X_t)\Big|\mathcal{Y}_t\Big] \ . \tag{2.4}$$

Let us define the measure $\widetilde{\mathbb{P}}$ by

$$\left.\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}\right|_{\mathcal{F}_t} = Z_t \tag{2.5}$$

and the associated inverse relation

$$\left.\frac{d\mathbb{P}}{d\widetilde{\mathbb{P}}}\right|_{\mathcal{F}_t} = \widetilde{Z}_t \tag{2.6}$$

where $\widetilde{Z}_t = Z_t^{-1}$ can be written as

$$\widetilde{Z}_t = \exp\left(\int_0^t h_s(X_s)^\top dY_s - \frac{1}{2}\int_0^t ||h(X_s)||^2 ds\right) \ . \tag{2.7}$$

Note that the process Y becomes a standard Brownian motion in the measure $\widetilde{\mathbb{P}}$.

We define the unnormalized conditional distribution of X to be the measure-valued process $\rho = \{\rho_t, t \geq 0\}$

$$\rho_t(\varphi) = \widetilde{\mathbb{E}}\Big[\widetilde{Z}_t\varphi(X_t)\Big|\mathcal{Y}_t\Big] \qquad \widetilde{\mathbb{P}} - a.s. \tag{2.8}$$

which is $\mathcal{Y}_t$-adapted and càdlàg. Here, $\widetilde{\mathbb{E}}[\cdot]$ denotes the expectation in the measure $\widetilde{\mathbb{P}}$. The desired filtered density function can then be obtained from the relation

$$\pi_t(\varphi) = \frac{\rho_t(\varphi)}{\rho_t(\mathbf{1})}. \tag{2.9}$$

It is known that the dynamics of ρ satisfies the following *Zakai* equation for arbitrary bounded function φ:

$$\rho_t(\varphi) = \rho_0(\varphi) + \int_0^t \rho_s(A_s\varphi)ds + \int_0^t \rho_s\Big((h_s + B_s)^\top\varphi\Big)dY_s \tag{2.10}$$

with initial value $\rho_0(\varphi) = \mathbb{E}[\varphi(X_0)]$ associated with a given distribution of X_0. Here, A_s is the infinitesimal generator of X at time s

$$A_s = \sum_{i=1}^n \mu_s^i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^n \big(\eta_s\eta_s^\top(x) + \bar{\eta}_s\bar{\eta}_s^\top(x)\big)_{i,j}\frac{\partial^2}{\partial x_i\partial x_j} \tag{2.11}$$

and

$$B_s^k = \sum_{i=1}^n (\bar{\eta}_s^\top(x))_{k,i}\frac{\partial}{\partial x_i}, \qquad k = 1, \cdots, m. \tag{2.12}$$

For the derivation of the Zakai equation and the other technical details, see [1], for example. The goal of this paper is to develop a simple scheme to solve the Zakai equation (2.10).

2.2 Filtered Characteristic Function

Let us consider a function

$$\psi(\xi, x) = \exp\Big(i\xi^\top x\Big) \tag{2.13}$$

with $\xi, x \in \mathbb{R}^n$, where $i = \sqrt{-1}$. If one obtains the conditional expectation of this function, i.e.,

$$\pi_t(\psi(\xi,\cdot)) = \frac{\widetilde{\mathbb{E}}\Big[\exp\Big(i\xi^\top X_t\Big)\widetilde{Z}_t\Big|\mathcal{Y}_t\Big]}{\widetilde{\mathbb{E}}\big[\widetilde{Z}_t\big|\mathcal{Y}_t\big]} \tag{2.14}$$

for each ξ, it enables one to derive the conditional expectation for an arbitrary choice of φ. This fact can be seen as follows: Let us consider the inverse Fourier transformation $\phi_t(\cdot)$

$$\phi_t(z) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} e^{-i\xi^\top z}\pi_t(\psi(\xi,\cdot))d^n\xi \tag{2.15}$$

which can be evaluated as

$$\begin{aligned}\phi_t(z) &= \frac{1}{(2\pi)^n}\int_{\mathbb{R}^d}\frac{1}{\widetilde{\mathbb{E}}\big[\widetilde{Z}_t\big|\mathcal{Y}_t\big]}\widetilde{\mathbb{E}}\Big[\exp\Big(i\xi^\top(X_t - z)\Big)\widetilde{Z}_t\Big|\mathcal{Y}_t\Big]d^n\xi \\ &= \frac{\widetilde{\mathbb{E}}\Big[\delta(X_t - z)\widetilde{Z}_t\Big|\mathcal{Y}_t\Big]}{\widetilde{\mathbb{E}}\big[\widetilde{Z}_t\big|\mathcal{Y}_t\big]}\ .\end{aligned} \tag{2.16}$$

The above function actually corresponds to the conditional density of the X_t since

$$\int_{\mathbb{R}^d}\varphi(z)\phi_t(z)d^nz = \frac{\widetilde{\mathbb{E}}\Big[\varphi(X_t)\widetilde{Z}_t\Big|\mathcal{Y}_t\Big]}{\widetilde{\mathbb{E}}\big[\widetilde{Z}_t\big|\mathcal{Y}_t\big]} \tag{2.17}$$

$$= \pi_t(\varphi)\ . \tag{2.18}$$

Therefore, $\{\pi_t(\psi(\xi))\}$ and hence $\{\rho_t(\psi(\xi))\}$ contains all the important information one needs.

3 Asymptotic Expansion in Momentum Space

From the discussion in the previous section, one needs to solve the Zakai equation for $\rho_t(\psi(\xi))$. However, the equation looks quite complicated because of the nonlinearities arising from the terms $\rho_s(A_s\psi(\xi))$ and $\rho_s((h_s + B_s)^\top\psi(\xi))$.

3.1 Perturbed System

In order to make the system tractable, we introduce the perturbation parameter ϵ and consider the following n-dimensional signal $X^{(\epsilon)}$ and the m-dimensional observation Y processes:

$$dX_t^{(\epsilon)} = \Big(f_t + \epsilon F_t(X_t^{(\epsilon)})\Big)dt + \Big(\nu_t + \epsilon\sigma_t(X_t^{(\epsilon)})\Big)dV_t + \epsilon\gamma_t(X_t^{(\epsilon)})dW_t \tag{3.1}$$

$$dY_t = \epsilon H_t(X_t^{(\epsilon)})dt + dW_t \tag{3.2}$$

where now f_t and ν_t are n-dimensional deterministic function of time and $\nu\nu_t^\top$ is assumed to be positive definite [2]. As before, we assume $Y_0 = 0$ and the initial condition of $X_0^{(\epsilon)} = X_0$ is known and given by an independent distribution. $F_t(x), \sigma_t(x), \gamma_t(x)$ and $H_t(x)$ take values in $\mathbb{R}^n$, $\mathbb{R}^{n\times d}$, $\mathbb{R}^{n\times m}$ and $\mathbb{R}^m$, respectively. Furthermore, these functions are assumed to be given by smooth and bounded functions to guarantee the existence of unique solution of $X^{(\epsilon)}$. As before, V and W are independent Brownian motion with dimension d and m, respectively.

As explained in the previous section, we are interested in the unnormalized distribution

$$\rho^{(\epsilon)}(\psi(\xi)) = \widetilde{\mathbb{E}}\Big[\widetilde{Z}_t^{(\epsilon)}\psi(\xi, X_t^{(\epsilon)})\Big|\mathcal{Y}_t\Big] \tag{3.3}$$

with

$$\widetilde{Z}_t^{(\epsilon)} = \exp\left(\epsilon\int_0^t H_s(X_s^{(\epsilon)})^\top dY_s - \frac{\epsilon^2}{2}\int_0^t ||H_s(X_s^{(\epsilon)})||^2 ds\right) \tag{3.4}$$

and

$$\left.\frac{d\mathbb{P}}{d\widetilde{\mathbb{P}}}\right|_{\mathcal{F}_t} = \widetilde{Z}_t^{(\epsilon)} \ . \tag{3.5}$$

The corresponding Zakai equation is now becomes

$$\rho_t^{(\epsilon)}(\psi(\xi)) = \rho_0(\psi(\xi)) + \int_0^t \rho_s^{(\epsilon)}\Big(A_s^{(\epsilon)}\psi(\xi)\Big)ds + \epsilon\int_0^t \rho_s^{(\epsilon)}\Big((H_s + B_s)^\top\psi(\xi)\Big)dY_s \ . \tag{3.6}$$

Here,

$$B_s^k = \sum_{i=1}^n (\gamma_s^\top(x))_{k,i}\frac{\partial}{\partial x_i}, \qquad k = 1, \cdots, m. \tag{3.7}$$

and the infinitesimal generator is given by

$$\begin{aligned} A_s^{(\epsilon)} &= \sum_{i=1}^n\Big(f_s^i + \epsilon F_s^i(x)\Big)\frac{\partial}{\partial x_i} + \sum_{i,j=1}^n \frac{1}{2}\Big(\nu_s + \epsilon\sigma_s(x)\Big)\Big(\nu_s + \epsilon\sigma_s(x)\Big)_{ij}^\top \frac{\partial^2}{\partial x_i \partial x_j} \\ &\quad + \sum_{i,j=1}^n \epsilon^2\frac{1}{2}\Big(\gamma_s(x)\gamma_s(x)^\top\Big)_{ij}\frac{\partial^2}{\partial x_i\partial x_j} \ . \end{aligned} \tag{3.8}$$

Our goal is to expand

$$\rho_t^{(\epsilon)}(\psi(\xi)) = \rho_t^{(0)}(\psi(\xi)) + \epsilon\rho_t^{(1)}(\psi(\xi)) + \epsilon^2\rho_t^{(2)}(\psi(\xi)) + \cdots \tag{3.9}$$

and obtain each $\rho_t^{(j)}(\psi(\xi))$ for $j = 0, 1, 2\cdots$.

The above asymptotic expansion gives a good approximation of the original model only when the perturbation terms are small enough, otherwise the above expansion fails to converge. However, as we shall see later, a simple *substepping method* for asymptotic expansion can make the restriction significantly weaker.

[2] We do not put ϵ on Y since it is observed process.

3.2 Asymptotic Expansion

We now expand the Zakai equation for each order of ϵ. Note that, for any polynomial function G of x, one can write

$$G(x)e^{i\xi^\top x} = G(D_\xi)e^{i\xi^\top x} \tag{3.10}$$

where, $G(D_\xi)$ denotes the differential operator obtained by replacing each x_j in the function by $(D_\xi)_j$, which is a derivative operator defined by

$$D_\xi = \frac{\partial}{i\partial\xi} \ . \tag{3.11}$$

This fact allows to write

$$\rho_t^{(\epsilon)}(G\psi(\xi)) = G(D_\xi)\rho_t^{(\epsilon)}(\psi(\xi)) \tag{3.12}$$

which is linear for $\rho_t^{(\epsilon)}(\psi(\xi))$.

In order to avoid nonlinearity, we make use of this property of polynomial functions. With slight abuse of notations, we treat $F_t(x), \sigma_t(x), \gamma_t(x)$ and $H_t(x)$ as arbitrary accurately approximated polynomial functions of x (and time) for the corresponding original functions. By Weierstrass' polynomial approximation theorem, this is always possible for any continuous functions within the closed interval. In practice, one can take wide enough interval within which the signal process resides with probability sufficiently close to one and an associated polynomial approximation accurate enough for that range.

Then, one can formally write

$$A_s^{(\epsilon)}\psi(\xi,x) = \Big(A_s^{(0)}(\xi) + \epsilon A_s^{(1)}(\xi, D_\xi) + \epsilon^2 A_t^{(2)}(\xi, D_\xi)\Big)\psi(\xi,x) \tag{3.13}$$

where

$$A_s^{(0)}(\xi) = i\xi^\top f_s - \frac{1}{2}\xi^\top(\nu_s\nu_s^\top)\xi \tag{3.14}$$

$$A_s^{(1)}(\xi, D_\xi) = i\xi^\top F_s(D_\xi) - \frac{1}{2}\xi^\top\Big(\nu_s\sigma_s^\top(D_\xi) + \sigma_s(D_\xi)\nu_s^\top\Big)\xi \tag{3.15}$$

$$A_s^{(2)}(\xi, D_\xi) = -\frac{1}{2}\xi^\top\Big(\sigma_s(D_\xi)\sigma_s^\top(D_\xi) + \gamma_s(D_\xi)\gamma_s^\top(D_\xi)\Big)\xi \tag{3.16}$$

and similarly

$$(H_s(x) + B_s(x))^\top\psi(\xi,x) = \Big(H_s^\top(D_\xi) + i\xi^\top\gamma_s(D_\xi)\Big)\psi(\xi,x) \ . \tag{3.17}$$

These observations lead to the following result:

Theorem 1 *Each order of the asymptotic expansion $\rho_t^{(j)}(\psi(\xi))$ in*

$$\rho_t^{(\epsilon)}(\psi(\xi)) = \rho_t^{(0)}(\psi(\xi)) + \epsilon\rho_t^{(1)}(\psi(\xi)) + \epsilon^2\rho_t^{(2)}(\psi(\xi)) + \cdots \tag{3.18}$$

of the unnormalized filtered characteristic function

$$\rho_t^{(\epsilon)}(\psi(\xi)) = \widetilde{\mathbb{E}}\Big[\exp\Big(i\xi^\top X_t^{(\epsilon)}\Big)\widetilde{Z}_t^{(\epsilon)}\Big|\mathcal{Y}_t\Big] \tag{3.19}$$

satisfies

$$\begin{aligned} d\rho_t^{(j)}(\psi(\xi)) &= A_t^{(0)}(\xi)\rho_t^{(j)}(\psi(\xi))dt \\ &+\Big\{A_t^{(1)}(\xi, D_\xi)\rho_t^{(j-1)}(\psi(\xi)) + A_t^{(2)}(\xi, D_\xi)\rho_t^{(j-2)}(\psi(\xi))\Big\}dt \\ &+\Big(H_t^\top(D_\xi) + i\xi^\top\gamma_t(D_\xi)\Big)\rho_t^{(j-1)}(\psi(\xi))dY_t \end{aligned} \tag{3.20}$$

with initial condition $\rho_0^{(0)}(\psi(\xi)) = \rho_0(\psi(\xi))$ *and the convention that*

$$\rho^{(j)}(\psi(\xi)) \equiv 0 \tag{3.21}$$

for $j < 0$.

Considering a special case where there is no observation, one obtains a simple corollary for a standard unconditional characteristic function.

Corollary 1 *Each order of the asymptotic expansion* $\rho_t^{(j)}(\psi(\xi))$ *in*

$$\rho_t^{(\epsilon)}(\psi(\xi)) = \rho_t^{(0)}(\psi(\xi)) + \epsilon\rho_t^{(1)}(\psi(\xi)) + \epsilon^2\rho_t^{(2)}(\psi(\xi)) + \cdots \tag{3.22}$$

of the characteristic function

$$\rho_t^{(\epsilon)}(\psi(\xi)) = \mathbb{E}\Big[\exp\Big(i\xi^\top X_t^{(\epsilon)}\Big)\Big] \tag{3.23}$$

satisfies

$$\begin{aligned} d\rho_t^{(j)}(\psi(\xi)) &= A_t^{(0)}(\xi)\rho_t^{(j)}(\psi(\xi))dt \\ &+\Big\{A_t^{(1)}(\xi, D_\xi)\rho_t^{(j-1)}(\psi(\xi)) + A_t^{(2)}(\xi, D_\xi)\rho_t^{(j-2)}(\psi(\xi))\Big\}dt \end{aligned}$$

with the convention that

$$\rho^{(j)}(\psi(\xi)) \equiv 0 \tag{3.24}$$

for $j < 0$.

The above result shows that one only has to deal with a set of decoupled ODEs in terms of momentum $\{\xi\}$ with a given observation path of Y. It is straightforward to solve the above equation for each ξ up to a certain ϵ-order, and use discrete Fourier transformation technique to obtain the density function. In Fourier analysis of smooth functions, it is well-known that most of the information is carried by small number of modes. In fact, in an example we provide in a later section, the resultant density function does not change meaningfully once the number of ξ-modes reaches ~ 30. This feature combined with the decoupled dynamics of characteristic function is expected to weaken the curse of dimensionality significantly, at least compared to typical PDE approaches.

Analytical calculation is also possible. Since the dynamics is linear, one easily obtains the following results:

Zeroth order

$$\rho_t^{(0)}(\psi(\xi)) = e^{\int_0^t A_s^{(0)}(\xi)ds}\rho_0(\psi(\xi)) \tag{3.25}$$

First order

$$\rho_t^{(1)}(\psi(\xi)) = \int_0^t e^{\int_s^t A_u^{(0)}(\xi)du}\Big\{A_s^{(1)}(\xi, D_\xi)\rho_s^{(0)}(\psi(\xi))ds + \Big(H_s^\top(D_\xi) + i\xi^\top\gamma_s(D_\xi)\Big)\rho_s^{(0)}(\psi(\xi))dY_s\Big\} \tag{3.26}$$

Higher order ($j \geq 2$)

$$\rho_t^{(j)}(\psi(\xi)) = \int_0^t e^{\int_s^t A_u^{(0)}(\xi)du}\Big\{\Big(A_s^{(1)}(\xi, D_\xi)\rho_s^{(j-1)}(\psi(\xi)) + A_s^{(2)}(\xi, D_\xi)\rho_s^{(j-2)}(\psi(\xi))\Big)ds + \Big(H_s^\top(D_\xi) + i\xi^\top\gamma_s(D_\xi)\Big)\rho_s^{(j-1)}(\psi(\xi))dY_s\Big\} \ . \tag{3.27}$$

In the next section, we explain the inversion method to obtain the density function.

3.3 Density Formula

Let us consider the inverse Fourier transformation of $\rho_t^{(\epsilon)}(\psi(\xi))$:

$$\phi_t^{(\epsilon)}(z) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} e^{-i\xi^\top z}\rho_t^{(\epsilon)}(\psi(\xi))d^n\xi \ . \tag{3.28}$$

This corresponds to the unnormalized conditional probability density of the signal $X_t^{(\epsilon)}$ given the observation path of Y (See the discussion in Sec.2.2.). The desired normalized conditional probability density of the signal is then given by

$$\bar{\phi}_t^{(\epsilon)}(z) = \frac{1}{c_t^{(\epsilon)}}\phi_t^{(\epsilon)}(z) \tag{3.29}$$

where $c_t^{(\epsilon)}$ is a normalization factor defined as

$$\Big(c_t^{(\epsilon)}\Big)^{-1} = \int_{\mathbb{R}^n}\phi_t^{(\epsilon)}(z)d^nz \ . \tag{3.30}$$

Thus, for applications, it suffices to calculate the expression of $\phi_t^{(\epsilon)}(z)$.

3.3.1 Gaussian distribution for X_0

For simplicity, let us first consider the case where X_0 is distributed by a Gaussian law $\mathcal{N}(x_0;\Sigma_0)$ with mean x_0 and the covariance Σ_0 of symmetric positive definite matrix. In this case, we have

$$\rho_0(\psi(\xi)) = \int_{\mathbb{R}^n} e^{i\xi^\top x}n[x;x_0,\Sigma_0]d^nx \tag{3.31}$$

$$= \frac{1}{\sqrt{(2\pi)^n|\Sigma_0|}}\int_{\mathbb{R}^n} e^{i\xi^\top x}\exp\left(-\frac{1}{2}(x-x_0)^\top\Sigma_0^{-1}(x-x_0)\right)d^nx \tag{3.32}$$

where $n[x; x_0, \Sigma_0]$ is the probability density function for a random variable with Gaussian low of $\mathcal{N}(x_0; \Sigma_0)$, and $|\Sigma_0|$ denotes the determinant of Σ_0. The evaluation can be done easily by considering the variable change from x to η given by

$$x = x_0 + P_0\eta \tag{3.33}$$

with a matrix P_0 satisfying

$$\Sigma_0 = P_0 P_0^\top. \tag{3.34}$$

Integration in terms of η easily leads to

$$\rho_0(\psi(\xi)) = \exp\Big(i\xi^\top x_0 - \frac{1}{2}\xi^\top \Sigma_0 \xi\Big) \ . \tag{3.35}$$

Then, from the result of previous section, we have

$$\rho_t^{(0)} = \exp\Big(i\xi^\top x_t - \frac{1}{2}\xi^\top \Sigma_t \xi\Big) \tag{3.36}$$

where

$$x_t = x_0 + \int_0^t f_s ds \tag{3.37}$$

$$\Sigma_t = \Sigma_0 + \int_0^t \nu_s \nu_s^\top ds \ . \tag{3.38}$$

Thus, it is clear that $X_t^{(0)}$ has a Gaussian distribution $\mathcal{N}(x_t; \Sigma_t)$. If the initial position of X_0 is exactly known as $X_0 = x_0$, then one clearly has

$$\rho_0(\psi(\xi)) = e^{i\xi^\top x_0} \tag{3.39}$$

and hence one can simply insert $\Sigma_0 = 0$ in (3.36).

By the property of the exponential form and (3.26), one can check that $\rho_t^{(1)}(\psi(\xi))$ is given by

$$\rho_t^{(1)}(\psi(\xi)) = \rho_t^{(0)}(\psi(\xi)) \left(\int_0^t a_s(\xi) ds + b_s(\xi) dY_s \right) \tag{3.40}$$

with polynomial functions $a_s(\xi) \in \mathbb{R}$ and $b_s(\xi) \in \mathbb{R}^{1\times m}$ of ξ

$$a_s(\xi) = \rho_s^{(0)}(\psi(\xi))^{-1} A_s^{(1)}(\xi, D_\xi) \rho_s^{(0)}(\psi(\xi)) \tag{3.41}$$

$$b_s(\xi) = \rho_s^{(0)}(\psi(\xi))^{-1} \Big(H_s(D_\xi)^\top + i\xi^\top \gamma_s(D_\xi) \Big) \rho_s^{(0)}(\psi(\xi)) \tag{3.42}$$

which can be expressed by Hermite polynomials in general.

Thus, the first order correction to the unnormalized conditional density can be expressed as

$$\phi_t^{(1)}(z) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-i\xi^\top z} \rho_t^{(1)}(\psi(\xi)) d^n\xi \tag{3.43}$$

$$= \left(\int_0^t a_s(D_z) ds + b_s(D_z)^\top dY_s \right) n[z; x_t, \Sigma_t] \ . \tag{3.44}$$

Here,

$$D_z = i\frac{\partial}{\partial z} \tag{3.45}$$

and $a_s(D_z)$, $b_s(D_z)$ denote the derivative operator of z obtained by replacing each ξ in the functions by D_z.

Repeating the same arguments, one can confirm that $\rho_t^{(j)}(\psi(\xi))$ can be written, in general, as

$$\begin{aligned}\rho_t^{(1)}(\psi(\xi)) &= \rho_t^{(0)}(\psi(\xi))\int_0^t\int_0^{s_j}\cdots\int_0^{s_2}\Big(\Gamma_{s_1,\cdots,s_j}(\xi)ds_1ds_2\cdots ds_j \\ &\quad +\Gamma_{y_1,s_2,\cdots,s_j}(\xi)\ dY_{s_1}ds_2\cdots ds_j \\ &\quad \cdots \\ &\quad +\Gamma_{y_1,y_2,\cdots,y_j}(\xi)\ dY_{s_1}dY_{s_2}\cdots dY_{s_j}\Big)+\cdots \end{aligned} \tag{3.46}$$

with certain polynomial functions $\{\Gamma_{s_1,\cdots,s_j}(\xi),\cdots,\Gamma_{y_1,\cdots,y_j}(\xi)\}$ of ξ with appropriate dimensions. The $\{\cdots\}$ denotes the term with integration order of $(<j)$, which stems from the existence of the ϵ-second order operator $A^{(2)}$. Therefore, the j-th order term of the unnormalized conditional density is also given by the correction to the Gaussian distribution:

$$\phi_t^{(j)}(z) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} e^{-i\xi^\top z}\rho_t^{(j)}(\psi(\xi))d^n\xi \tag{3.47}$$

$$\begin{aligned} &= \int_0^t\int_0^{s_j}\cdots\int_0^{s_2}\Big(\Gamma_{s_1,\cdots,s_j}(D_z)ds_1ds_2\cdots ds_j \\ &\quad +\Gamma_{y_1,s_2,\cdots,s_j}(D_z)\ dY_{s_1}ds_2\cdots ds_j \\ &\quad \cdots \\ &\quad +\Gamma_{y_1,y_2,\cdots,y_j}(D_z)\ dY_{s_1}dY_{s_2}\cdots dY_{s_j}\Big)n[z;x_t,\Sigma_t]\ +\cdots . \end{aligned} \tag{3.48}$$

In the case of no (or trivial) observation, one can get the asymptotic expansion of unconditional probability density by putting dY terms zero.

3.3.2 Non-Gaussian distribution for X_0

Even when the initial distribution is not exactly Gaussian, if one can approximate it by the form

$$\phi_0(z) = (\text{some polynomial function of z})\times n[z;x_0,\Sigma_0]\ , \tag{3.49}$$

then the properties of the inverse transformation given in the previous section still hold in almost the same way. This is, for example, the case when one approximates the initial distribution by Gram-Charlier expansions. In the case when (3.49) holds, one can still write $\rho_t^{(0)}$ in the form

$$\rho_t^{(0)}(\psi(\xi)) = (\text{some polynomial function of } \xi)\times\exp\Big(i\xi^\top x_t - \frac{1}{2}\xi^\top\Sigma_t\xi\Big) \tag{3.50}$$

and it only changes the functions $\{\Gamma\}$ in (3.48).

4 A Direct Application to Kushner-Stratonovich Equation

We can also apply the technique to the Kushner-Stratonovich (KS) equation that describes the dynamics of the normalized conditional density of π_t instead of ρ_t. Although it suffices to work on the simpler Zakai equation in filtering problems, we directly treat KS equation here to demonstrate the fact that the asymptotic expansion can also be applied to measure-valued non-linear systems. For the setup given in Sec.2.1, the Kushner-Stratonovich equation is given by

$$d\pi_t(\varphi) = \pi_t(A_t\varphi)dt + \Big\{\pi_t\Big((h_t + B_t)^\top\varphi\Big) - \pi_t(h_t^\top)\pi_t(\varphi)\Big\}(dY_t - \pi_t(h_t)dt) \tag{4.1}$$

with a given initial value $\pi_0(\varphi)$. This is clearly a nonlinear equation for the measure-valued process π_t. See a textbook [1] for details of the derivation.

Let us now introduce the same perturbed system as in Sec.3.1. Then, one obtains the KS equation for $\psi(\xi,\cdot)$ as

$$\begin{aligned} d\pi_t^{(\epsilon)}(\psi(\xi)) &= \pi_t^{(\epsilon)}(A_t^{(\epsilon)}\psi(\xi))dt \\ &\quad +\epsilon\Big\{\pi_t^{(\epsilon)}\Big((H_t + B_t)^\top\psi(\xi)\Big) - \pi_t^{(\epsilon)}(H_t^\top)\pi_t^{(\epsilon)}(\psi(\xi))\Big\}(dY_t - \epsilon\pi_t^{(\epsilon)}(H_t)dt) \end{aligned} \tag{4.2}$$

By the same polynomial-function approximations, one can rewrite it as

$$\begin{aligned} d\pi_t^{(\epsilon)}(\psi(\xi)) &= \Big(A_t^{(0)}(\xi) + \epsilon A_t^{(1)}(\xi, D_\xi) + \epsilon^2 A_t^{(2)}(\xi, D_\xi)\Big)\pi_t^{(\epsilon)}(\psi(\xi))dt \\ &\quad +\epsilon\Big\{\Big(H_t^\top(D_\xi) + i\xi^\top\gamma_t(D_\xi)\Big)\pi_t^{(\epsilon)}(\psi(\xi)) - \pi_t^{(\epsilon)}(H_t^\top)\pi_t^{(\epsilon)}(\psi(\xi))\Big\}(dY_t - \epsilon\pi_t^{(\epsilon)}(H_t)dt)\ . \end{aligned} \tag{4.3}$$

Now, there appears $\pi_t^{(\epsilon)}(H_t)$ in the equation. This term does no harm since an arbitrary order j of asymptotic expansion, we need $\pi_t^{(i)}(H_t)$ only for $i = 0, 1, \cdots, j-1$ due to the additional ϵ-factor. Thus, at the calculation of j-th order expansion, one can use

$$\pi_t^{(i)}(H_t) = H_t(D_\xi)\pi_t^{(i)}(\psi(\xi))\Big|_{\xi=0} \tag{4.4}$$

where $\pi_t^{(i)}(\psi(\xi))$ are already known for $i \le j-1$.

Let us give the first few orders of expansions for the KS equation:
Zeroth order

$$d\pi_t^{(0)}(\psi(\xi)) = A_t^{(0)}(\xi)\pi_t^{(0)}(\psi(\xi))dt \tag{4.5}$$

with initial value $\pi_0^{(0)}(\psi(\xi)) = \pi_0(\psi(\xi))$.
First oder

$$\begin{aligned} d\pi_t^{(1)}(\psi(\xi)) &= A_t^{(0)}(\xi)\pi_t^{(1)}(\psi(\xi))dt \\ &\quad +\Big\{\Big(H_t^\top(D_\xi) + i\xi^\top\gamma_t(D_\xi)\Big)\pi_t^{(0)}(\psi(\xi)) - \pi_t^{(0)}(H_t^\top)\pi_t^{(0)}(\psi(\xi))\Big\}dY_t \end{aligned} \tag{4.6}$$

with $\pi_0^{(1)}(\psi(\xi)) = 0$.
Second order

$$
\begin{aligned}
d\pi_t^{(2)}(\psi(\xi)) &= A_t^{(0)}(\xi, D_\xi)\pi_t^{(2)}(\psi(\xi))dt \\
&+\Big\{ A_t^{(1)}(\xi, D_\xi)\pi_t^{(1)}(\psi(\xi)) + A_t^{(2)}(\xi, D_\xi)\pi_t^{(0)}(\psi(\xi)) \\
&\quad - \Big(H_t^\top(D_\xi) + i\xi^\top\gamma_t(D_\xi) - \pi_t^{(0)}(H_t^\top)\Big)\pi_t^{(0)}(\psi(\xi))\pi_t^{(0)}(H_t)\Big\}dt \\
&+\Big\{ \Big(H_t^\top(D_\xi) + i\xi^\top\gamma_t(D_\xi) - \pi_t^{(0)}(H_t^\top)\Big)\pi_t^{(1)}(\psi(\xi)) - \pi_t^{(1)}(H_t^\top)\pi_t^{(0)}(\psi(\xi))\Big\}dY_t
\end{aligned}
\tag{4.7}
$$

with initial condition $\pi_t^{(2)}(\psi(\xi)) = 0$.

Although the system of ODEs does not keep the same structure as the unnormalized distribution ρ, it is clear that one can still perform the perturbation order by order. Furthermore, from the discussion given in the next section, it is not always necessary to derive higher order asymptotic expansion for accurate estimation.

5 Substepping Method for Asymptotic Expansion

It is obvious by construction that the accuracy of approximation deteriorates once the cumulative contributions from perturbation terms

$$
\epsilon F, \quad \epsilon\sigma, \quad \epsilon\gamma, \quad \epsilon H \tag{5.1}
$$

become bigger. This is a common problem of asymptotic expansion methods for various applications. In particular for the filtering problems, requiring small perturbation terms seems rather restrictive since it indicates that one can treat only noisy observations (i.e. small H). In financial applications of the position-space asymptotic expansion, it is known that one needs higher order approximation to reach enough accuracy for practical use in long-dated or high-volatility situations. There exists many efforts to obtain higher order corrections systematically to tackle these problems. See for examples [4, 6, 7] for recent developments in this direction.

Let us consider the problem in the momentum-space approach. In Theorem 1, we have seen that the relation

$$
\begin{aligned}
d\rho_t^{(j)}(\psi(\xi)) &= A_t^{(0)}(\xi)\rho_t^{(j)}(\psi(\xi))dt \\
&\quad +\Big\{ A_t^{(1)}(\xi, D_\xi)\rho_t^{(j-1)}(\psi(\xi)) + A_t^{(2)}(\xi, D_\xi)\rho_t^{(j-2)}(\psi(\xi))\Big\}dt \\
&\quad +\Big(H_t^\top(D_\xi) + i\xi^\top\gamma_t(D_\xi)\Big)\rho_t^{(j-1)}(\psi(\xi))dY_t
\end{aligned}
\tag{5.2}
$$

determines the correction terms with a given initial condition $\rho_0(\psi(\xi))$. Although higher order calculation is straightforward, there exists simpler and more efficient way to improve the approximation. An obvious but important feature of (5.2) is the fact that it can be applied to an arbitrary initial condition $\rho_0(\psi(\xi))$. Since asymptotic expansion typically works very well for short maturities, the above feature naturally leads to the following

substepping idea for asymptotic expansion.

(1) Create an appropriate time grid $\{0 = T_0 < T_1 < \cdots < T_N = t\}$ *in such a way that the asymptotic expansion converges well within each sub-period* $[T_{i-1}, T_i]$.
(2) Solve (5.2) of each ξ *for* $s \in [T_0, T_1]$ *up to* k*-th orders of asymptotic expansions. This can be first or second order if the stepping size is small enough.*
(3) Update the initial condition for the next period $[T_1, T_2]$ *by setting*

$$\rho^{(0)}_{T_1}(\psi(\xi)) \leftarrow \left(\sum_{j=0}^{k} \epsilon^j \rho^{(j)}_{T_1}(\psi(\xi)) \right) \text{ obtained in step(2) .} \tag{5.3}$$

(4) Solve (5.2) for $[T_1, T_2]$ *with the updated initial condition.*
(5) Repeat the procedures till the final period to obtain $\rho^{(j)}_{T_N}(\psi(\xi))$.

Although performing the above method analytically by hand is quite laborious due to a large number of derivative operations, one can do it quite efficiently in numerical implementations. This is because amount of procedures required for dt and dY integration does not change in the above operations. In fact, one obtains accurate results faster by performing finer substepping with lower-order approximation than performing higher-order approximation without substepping.

The substepping method may be very useful for general measure-valued nonlinear equations. Although it is tedious to obtain the asymptotic expansion for complicated dynamics, it is definitely possible for the first few orders as we have shown in the example given in the previous section. If the approximation works well, at least within a very short period, the above numerical procedures will extend the effective region for the asymptotic expansion. If it is applied to a standard unconditional characteristic function, it should also offer an efficient option pricing method, particularly for long-dated and high-volatile setups.

6 Examples

6.1 Analytical Application to CIR Process

Now, let us first consider the approximation of one-dimensional CIR process with no filtering issue, which helps to give a concrete image how analytical procedures work. We study

$$dX_t = \theta(\mu - X_t)dt + \sigma\sqrt{X_t}dV_t \tag{6.1}$$

with $X_0 = \mu$. All the parameters θ, μ and σ are positive constants satisfy $2\theta\mu > \sigma^2$. Then, the probability density of X is known to have a non-central chi-squared distribution.

For asymptotic expansion, we treat it as the following perturbed system: [3]

$$dX^{(\epsilon)}_t = \epsilon\theta F(X^{(\epsilon)}_t)dt + \sigma\Big(\sqrt{\mu} + \epsilon R(X^{(\epsilon)}_t)\Big)dV_t \tag{6.2}$$

[3] One needs to put $\epsilon = 1$ at the end for the comparison to the original model.

with $X_0^{(\epsilon)} = \mu$. Here, we have defined

$$F(x) = \mu - x \tag{6.3}$$
$$R(x) = \text{Taylor expansion at } (x = \mu) \text{ of } (\sqrt{x} - \sqrt{\mu}) \ . \tag{6.4}$$

In this example, we are going to adopt the 3rd-order expansion for $R(x)$. Note that, Taylor expansion provides a good polynomial approximation only when the process X resides near μ. If volatility is very high, it may be better to perform a different method, such as the minimization of the least square differences for appropriate range. We shall see some examples in the next section. Systematic strategy for determining the optimal choice of polynomial function remains as an important future work.

Then, the infinitesimal generator is given by

$$A^{(\epsilon)} = \epsilon\theta F(x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2\Big(\sqrt{\mu} + \epsilon R(x)\Big)^2\frac{\partial^2}{\partial x^2} \tag{6.5}$$

and hence

$$A^{(0)}(\xi) = -\frac{1}{2}\xi^2\sigma^2\mu \tag{6.6}$$
$$A^{(1)}(\xi, D_\xi) = i\xi\theta F(D_\xi) - \xi^2\sigma^2\sqrt{\mu}R(D_\xi) \tag{6.7}$$
$$A^{(2)}(\xi, D_\xi) = -\frac{1}{2}\xi^2\sigma^2R^2(D_\xi) \ . \tag{6.8}$$

From Corollary 1, analytical calculation can be performed as follows:

6.1.1 Zeroth order

We have

$$d\rho_t^{(0)}(\psi(\xi)) = A^{(0)}(\xi)\rho_t^{(0)}(\psi(\xi))dt \tag{6.9}$$

with $\rho_0^{(0)}(\psi(\xi)) = e^{i\xi\mu}$. Thus, it gives

$$\rho_t^{(0)}(\psi(\xi)) = \exp\left(i\xi\mu - \frac{1}{2}\xi^2\sigma^2\mu t\right) \ . \tag{6.10}$$

Then, one obtains the zeroth order density function as

$$\phi_t^{(0)}(z) = \frac{1}{\sqrt{2\pi\Sigma_t}}\exp\left(-\frac{(\mu - z)^2}{2\Sigma_t}\right) \tag{6.11}$$

with the definition of $\Sigma_t = \mu\sigma^2 t$.

6.1.2 First order

The first order correction is given by

$$\rho_t^{(1)}(\psi(\xi)) = \int_0^t e^{(t-s)A^{(0)}(\xi)}A^{(1)}(\xi, D_\xi)\rho_s^{(0)}(\psi(\xi))ds \ . \tag{6.12}$$

By straightforward differential operations lead to

$$\begin{aligned}\Big(\rho_s^{(0)}(\psi(\xi))\Big)^{-1} & A^{(1)}(\xi, D_\xi)\rho_s^{(0)}(\psi(\xi)) \\ &= \frac{1}{8}s\sigma^2(8\theta\mu+\sigma^2)\xi^2 - \frac{1}{16}is\sigma^4(8\mu+3s\sigma^2)\xi^3 \\ &\quad -\frac{1}{8}s^2\mu\sigma^6\xi^4 + \frac{1}{16}is^3\mu\sigma^8\xi^5 \ . \end{aligned} \tag{6.13}$$

Thus, one has

$$\rho_t^{(1)}(\psi(\xi)) = \rho_t^{(0)}\Big(a_2\xi^2 + a_3\xi^3 + a_4\xi^4 + a_5\xi^5\Big) \tag{6.14}$$

and also

$$\begin{aligned}\phi_t^{(1)}(z) &= \frac{1}{2\pi}\int_{\mathbb{R}} e^{-i\xi z}\rho_t^{(1)}(\psi(\xi))d\xi \\ &= \left(-a_2\frac{\partial}{\partial z^2} - ia_3\frac{\partial^3}{\partial z^3} + a_4\frac{\partial^4}{\partial z^4} + ia_5\frac{\partial^5}{\partial z^5}\right)\phi_t^{(0)}(z)\end{aligned} \tag{6.15}$$

with the coefficients defined by

$$a_2 = \frac{1}{16}t^2\sigma^2(8\theta\mu+\sigma^2), \qquad a_3 = -\frac{1}{16}it^2\sigma^4(4\mu+t\sigma^2) \tag{6.16}$$

$$a_4 = -\frac{1}{24}t^3\mu\sigma^6, \qquad a_5 = \frac{1}{64}it^4\mu\sigma^8 \ . \tag{6.17}$$

Higher order calculation follows similarly with the help of analytical software if necessary. In Figs. 1 and 2, we have given numerical results up to 3rd order asymptotic expansions without substepping compared with the exact non-central chi-squared distribution. When volatility is large, there appear sizable deviation from the correct distribution for small x. This is understandable because Taylor expansion near the origin is not accurate. Except a neighbor of the origin and $(x < 0)$, one can see that the approximation well reproduces the desired density function.

6.2 A Numerical Application to Beneš Filter

Now, we study an example of Beneš filter, where the drift of the signal process is nonlinear. This is a special example for which there exists an exact solution in the class of non-Gaussian filtering problems, and hence quite useful to test the current method. In the class of Beneš filter [3], we choose a following one-dimensional example:

$$dX_t = f(X_t)dt + \sigma dV_t \tag{6.18}$$

$$dY_t = \Big(h_1X_t + h_2\Big)dt + dW_t \tag{6.19}$$

with $X_0 = Y_0 = 0$, where $f(x)$ is given by

$$f(x) = a\sigma\tanh\Big(a\frac{x}{\sigma}\Big) \ . \tag{6.20}$$

Here, a, σ, h_1 and h_2 are all constants. In this case, the exact filtered density of X_t is given by

$$\begin{aligned}\pi_t^{\text{exact}}(z) \\ = \frac{1}{n_t}\cosh\Big(a\frac{z}{\sigma}\Big)\exp\Big(-\frac{h_1}{2\sigma}\coth(th_1\sigma)z^2 \\ +\Big(h_1\int_0^t\frac{\sinh(sh_1\sigma)}{\sinh(th_1\sigma)}dY_s+\frac{h_2}{\sigma\sinh(th_1\sigma)}-\frac{h_2}{\sigma}\coth(th_1\sigma)\Big)z\Big)\end{aligned} \tag{6.21}$$

where n_t is the normalization factor to guarantee

$$\int_{\mathbb{R}}\pi_t^{\text{exact}}(z)dz = 1\ . \tag{6.22}$$

For this problem, we setup the following perturbed approximation: [4]

$$dX_t^{(\epsilon)} = \epsilon F(X_t^{(\epsilon)})dt + \sigma dV_t \tag{6.23}$$

$$dY_t = \epsilon H(X_t^{(\epsilon)})dt + dW_t \tag{6.24}$$

with $X_0^{(\epsilon)} = Y_0 = 0$. Here, we use

$$H(x) = h_1x + h_2 \tag{6.25}$$

$$F(x) = (\text{polynomial approximation of})\ f(x)\ . \tag{6.26}$$

We explain the details of polynomial approximation later.

The infinitesimal generator contains only up to ϵ-first order term. We have

$$A^{(\epsilon)} = \epsilon F(x)\frac{\partial}{\partial x}+\frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2} \tag{6.27}$$

and

$$A^{(0)}(\xi) = -\frac{1}{2}\xi^2\sigma^2 \tag{6.28}$$

$$A^{(1)}(\xi, D_\xi) = i\xi F(D_\xi) \tag{6.29}$$

From Theorem 1, one needs to solve

$$\rho_t^{(j)}(\psi(\xi)) = \int_0^t e^{-\frac{1}{2}(t-s)\xi^2\sigma^2}\Big\{i\xi F(D_\xi)\rho_s^{(j-1)}(\psi(\xi))ds+\Big(h_1D_\xi+h_2\Big)\rho_s^{(j-1)}(\psi(\xi))dY_s\Big\} \tag{6.30}$$

starting from the zeroth order solution:

$$\rho_t^{(0)}(\psi(\xi)) = \exp\Big(-\frac{1}{2}t\xi^2\sigma^2\Big)\ . \tag{6.31}$$

[4] As before, one needs to put $\epsilon = 1$ at the end for the comparison to the original model.

6.2.1 Polynomial-function approximation

We now discuss how to obtain the polynomial approximation $F(x)$ for the nonlinear drift of the signal:

$$f(x) = a\sigma \tanh(ax/\sigma) \ . \tag{6.32}$$

Due to the normalization by σ, we can roughly expect

$$\left| \frac{X_t}{\sigma} \right| \lesssim 1 \tag{6.33}$$

for $t \in [0, 1]$. This implies that Taylor expansion around $x = 0$ is a natural candidate of $F(x)$ when $|a| \lesssim 1$. When $|a| \gtrsim 1$, the two plateaus of $f(x)$ start to play an important role in the range (6.33). Unfortunately, however, Taylor expansion does not reproduce the plateaus but strongly diverging behavior within the range (6.33) instead, which destabilizes the numerical result. Thus, we take $[-5\sigma, 5\sigma]$ range with a step size of 0.2σ, and carry out least-square method (LSM) to fit a 11-dimensional odd function for $F(x)$. We also adopt the weight function $g(x)$ defined as

$$g(x) = \exp\Big(-w\frac{x^2}{2\sigma^2}\Big) \tag{6.34}$$

with various factor w. Here, $w = 0$ corresponds to a pure LSM in the 5σ range and the polynomial function well recovers the two plateaus of $f(x)$ in the wide range, while it has a relatively poor fit around the origin. On the other hand, higher w gives finer fit and hence finer description of the density near $x = 0$. In this case, however, if one continues to increase w it starts to destabilize the numerical result as in the case for Taylor expansion. Thus, we need to take a balance of this trade-off, especially when $|a| > 1$.

6.2.2 Numerical Results

In the following numerical examples, we take $t = 1$ as the maturity and use $1,000$ steps to create the sample observation (and signal) path. We then integrate (6.30) with the same time step $dt = 10^{-3}$ with a given path of Y. For differentiation, we use a standard finite difference method. Finally, a discrete Fourier transformation is used to obtain the density function.

In the first numerical example given in Fig. 3, we have used a set of parameters $\{a = 0.8, \sigma = 0.5, h_1 = 0.8, h_2 = 0.5\}$, and a sample path of Y given in the top graph. We have used $w = 2.0$ for getting coefficients of polynomial function $F(x)$. The middle graph for the conditional density functions contains the exact one denoted by a red line labeled as "Benes", estimated conditional densities from $(0th, 3rd, 20th)$-order asymptotic expansion without substepping method, and those from 1st order expansion with $(100, 1000)$ sub-periods. One can clearly see the benefit of substepping method explained in Sec. 5: Although there is no clear improvement from 3rd to 20th order approximation, the substepping method with small sub-period provides almost exact fit to the true density function.

In Fig. 4, we have used $h_1 = 10.0$, which is an example of *small-noise* observation. Since $a = 0.5$ is relatively small, polynomial approximation for $f(x)$ is quite accurate ($w = 2.0$). The calculation has been performed with substepping method with (100, 125, 200, 1,000) sub-periods. Fine substepping gives almost exact density even in this case. In

particular, the significant reduction of the variance of the conditional density due to high quality information provided by the observation process is well reproduced by the repeated application of asymptotic expansion. In this example, approximation without substepping is too unstable and gives only meaningless numerical results. As suggested in Sec. 5, the size of perturbation terms itself does not seem to be a relevant problem for asymptotic expansion *as long as* we have accurate enough polynomial function approximation and the substepping method.

When one increases a, $f(x)$ becomes like a step function and makes it difficult to achieve accurate polynomial approximation for the relevant range. Here, the choice of LSM weight w starts to affect the estimated density. In Fig. 5, we have studied the case of $a = 1.5$ with several choice of w. Here, all the calculations were done with substepping method with 1000 sub-periods. The estimated density are stable for $w = 0.5 \sim 2.5$, it starts to give completely meaningless results for higher w. Note that, the impacts of LSM weight is highly dependent on the order of the polynomial function. As is easily guessed, the change of estimated density is more significant when lower order polynomials are used. In Fig. 6, we have done similar analysis for an example of $a = 2.0$, which reveals more clearly separated two peaks.

Remark :
In the above examples, we have used time-independent function for $F(x)$. However, when there exists a significant drift for the signal process, such as large $|a|$ in the above example, making the polynomial functions time dependent is quite likely to provide better estimation. If we have the information about the evolution of the conditional mean, we can change the center of polynomial-function approximation to replicate the original nonlinear function more accurately in the relevant region. Initial guess can be obtained by extended Kalman-Bucy filter or time independent $F(x)$ in the current method, for examples.

7 Conclusions and Discussions

In the paper, we have developed an asymptotic expansion technique in momentum space. Fourier transformation combined with polynomial-function approximation gives a closed recursive system of ODEs as an asymptotic expansion for the unnormalized conditional characteristic function. Thanks to the simplicity of the ODE system, higher order calculation can be performed easily. It also allows an efficient implementation of substepping method of asymptotic expansion. As long as polynomial approximation of the nonlinear terms is accurate, the size of nonlinear terms ceases to be a big obstacle for obtaining an accurate estimation. Applications to more realistic multi-dimensional filtering problems as well as other (financial) problems, such as option pricing, are left for the future research.

Let us make a brief comment on the remaining problems and possible future directions of research. As one can see, the method still suffers from the curse of dimensionality although it should be significantly weaker than typical PDE approaches. In asymptotic expansion method in the position space, the problem of dimensionality does not exist. This is one clear advantage to work in the position space. On the other hand, the substepping method looks difficult to implement for the latter, which limits the applicable parameter space unless higher order correction terms converge well. It will be quite beneficial if there

is any way to combine the advantages from the both sides.

Since we can only use finite order polynomials, the estimated density depends how they are calibrated. Although one can check it by actually performing various numerical tests, it is desirable to have sound systematic strategy. Related to this problem, when one has nonlinear terms difficult to fit by polynomials, the idea of *change-of-variable* developed in [8] may be proved to be useful. Suppose, one defines a new process $\widetilde{X}$ by using some function $\Psi(\cdot)$ as

$$\widetilde{X}_t = \Psi(X_t) \ . \tag{7.1}$$

If $\widetilde{X}$ has drift and diffusion terms that are well approximated by polynomial functions, one can get more accurate estimation of

$$\widetilde{\rho}_t(\xi) = \mathbb{E}\Big[e^{i\xi^\top \widetilde{X}_t}\Big] \tag{7.2}$$

and hence also its density

$$\widetilde{\phi}_t(z) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} e^{-i\xi^\top z}\widetilde{\rho}_t(\xi)d^n\xi \ . \tag{7.3}$$

Then, one can recover the density of the original X_t as

$$\phi_t(z) = \widetilde{\phi}\Big(\Psi(z)\Big)\Big|J(z)\Big| \tag{7.4}$$

where $|J|$ denotes the determinant of a Jacobian matrix with the elements of $(\partial\Psi_i(z)/\partial z_j)$. Thus some of the errors can be absorbed if there exists an appropriate choice of Ψ. The research for this direction also looks worth pursuing.

Acknowledgment

The author thanks to Akihiko Takahashi of University of Tokyo for informative discussions to understand the connections to the asymptotic expansion in the position space.

References

[1] Bain, A., Crisan, D., (2008) "Fundamentals of Stochastic Filtering," Springer.

[2] Crisan, D., Rozovskiĭ, B., (2011) "The Oxford Handbook of Nonlinear Filtering," Oxford university press.

[3] Beneš, V.E., (1981) "Exact finite-dimensional filters for certain diffusions with nonlinear drift," Stochastics, 5(1-2):65-92.

[4] Li, C., (2010).“ Closed-form Expansion, Conditional Expectation, and Option Valuation, ” Guanghua School of Management, Peking University, Beijing, China.

[5] Takahashi, A., (1999), "An Asymptotic Expansion Approach to Pricing Contingent Claims, Asia-Pacific Financial Markets," Vol. 6, 115-151.

[6] Takahashi, A. and Takehara, K., (2007) “ Pricing Currency Options with a Market Model of Interest Rates under Jump-Diffusion Stochastic Volatility Processes of Spot Exchange Rates, ” Asia-Pacific Financial Markets, Vol.14 , pp. 69-121.

[7] Takahashi, A., Takehara, K. and Toda. M., (2010). ”A General Computation Scheme for a High-Order Asymptotic Expansion Method,” CARF Working Paper, F-272, forthcoming in International Journal of Theoretical and Applied Finance.

[8] Takahashi, A., Toda, M., (2012). ”Note on an Extension of an Asymptotic Expansion Scheme”, CARF Working Paper F-286, available at http://www.carf.e.u-tokyo.ac.jp/workingpaper/.

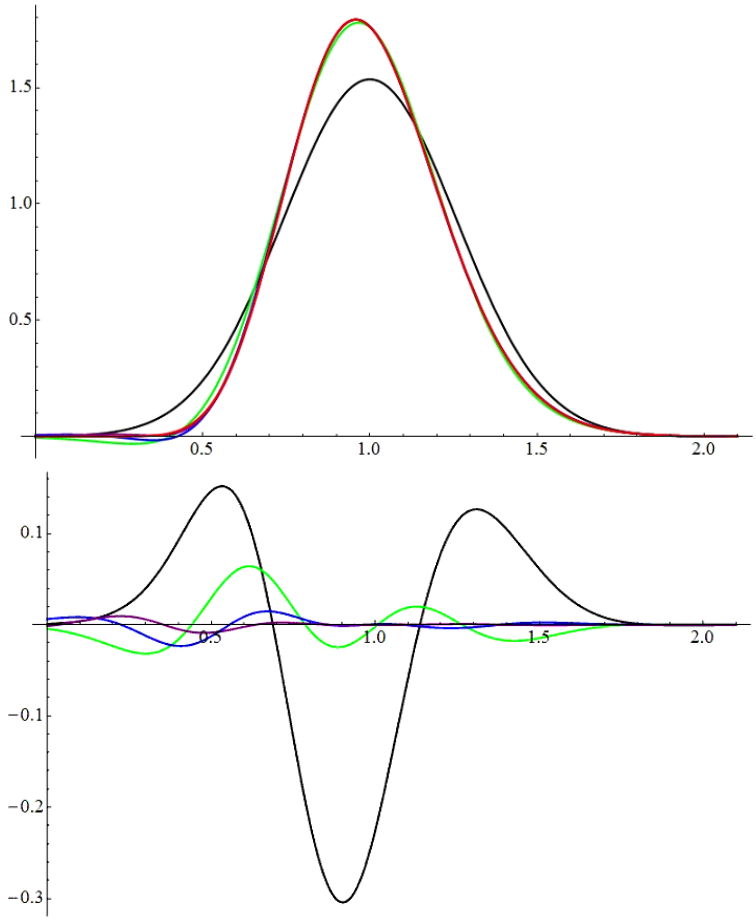

Figure 1: $t = 3, \mu = 1, \sigma = 0.15, \theta = 0.1$. (Black, Green, Blue, Purple) lines denote (0th, 1st, 2nd, 3rd) order approximation of asymptotic expansion, respectively. Red line denotes the exact density function given by a non-central chi-squared distribution. The second graph represents the difference from the exact density function.

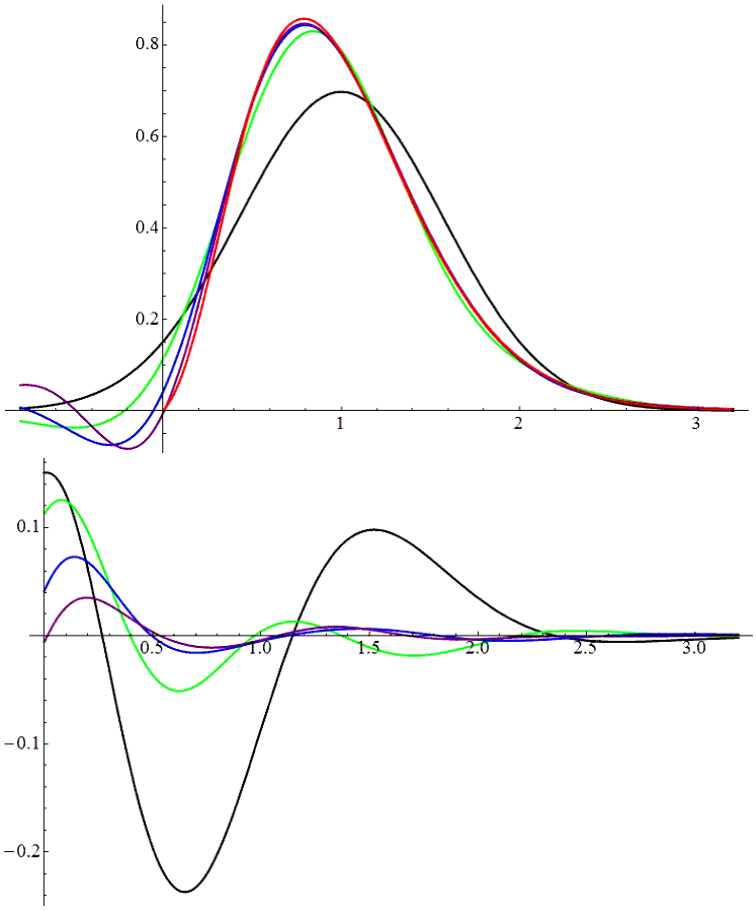

Figure 2: $t = 3, \mu = 1, \sigma = 0.33, \theta = 0.1$. (Black, Green, Blue, Purple) lines denote (0th, 1st, 2nd, 3rd) order approximation of asymptotic expansion, respectively. Red line denotes the exact density function given by a non-central chi-squared distribution. The second graph represents the difference from the exact density function.

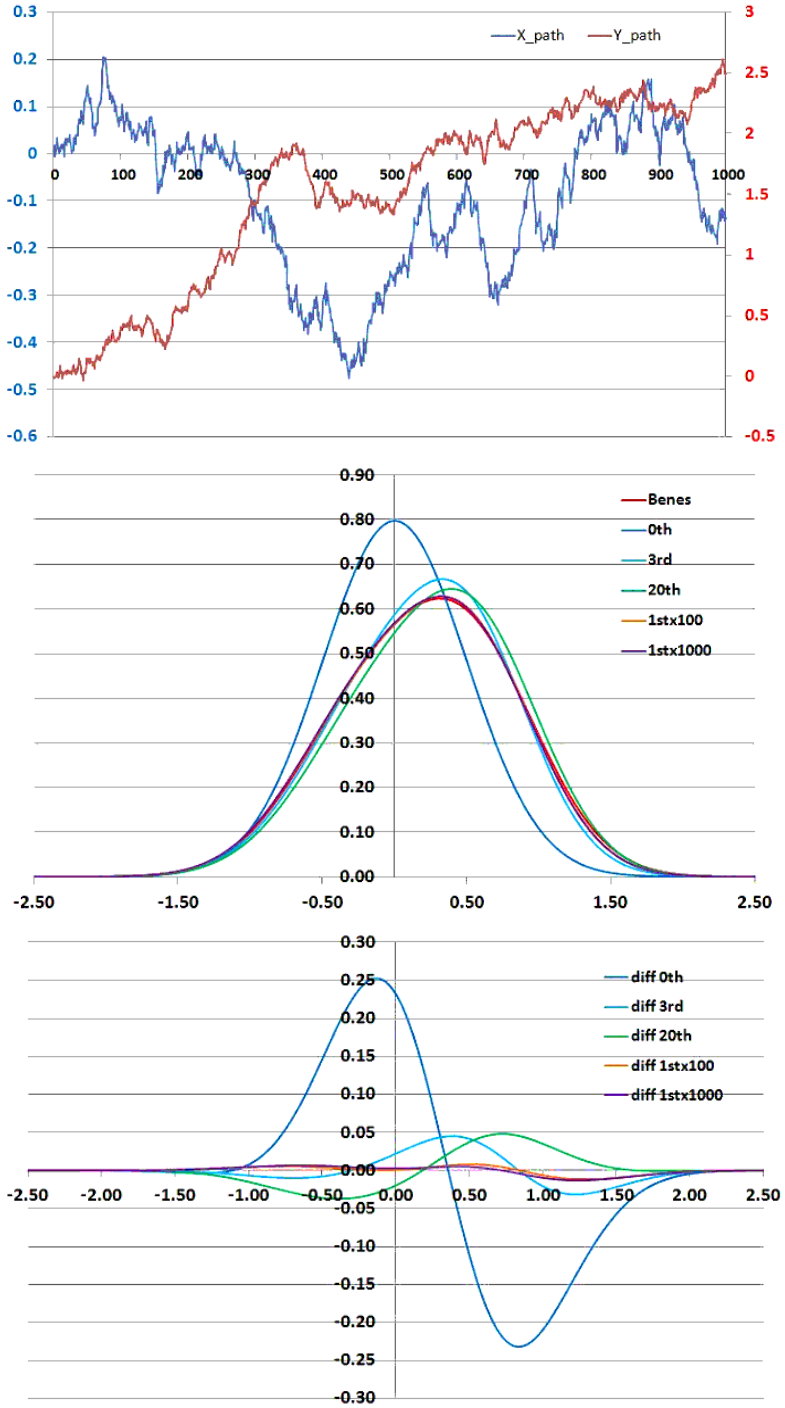

Figure 3: $t = 1, dt = 10^{-3}, a = 0.8, \sigma = 0.5, h_1 = 0.8, h_2 = 0.5$ with polynomial function fitted with $w = 2.0$. From top to bottom, the sample path, exact and approximated density functions, and the difference of the approximated densities from the exact one. In the middle graph, a red line labeled by "Benes" denotes the exact density function.

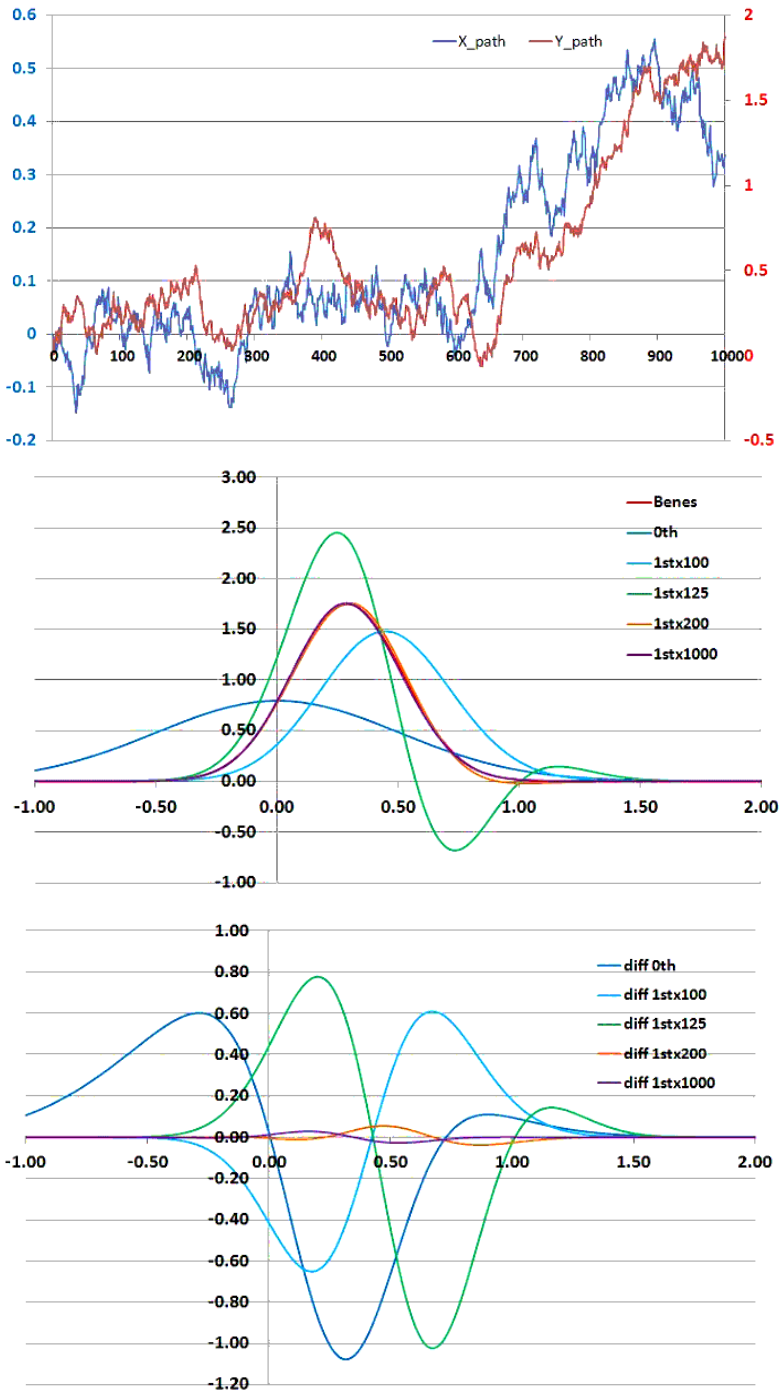

Figure 4: $t = 1, dt = 10^{-3}, a = 0.5, \sigma = 0.5, h_1 = 10.0, h_2 = 0.5$ with polynomial function fitted with $w = 2.0$. From top to bottom, the sample path, exact and approximated density functions, and the difference of the approximated densities from the exact one. In the middle graph, a red line labeled by "Benes" denotes the exact density function.

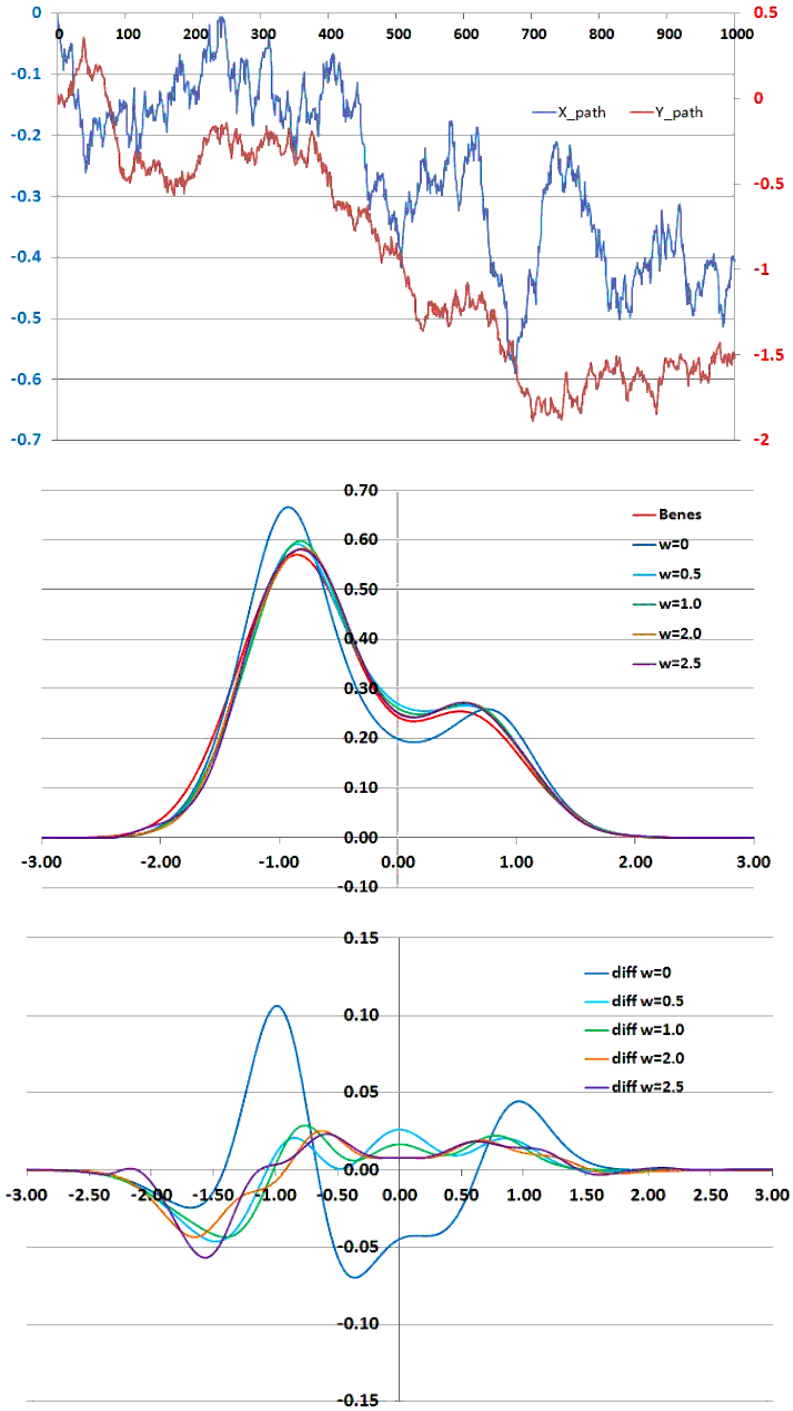

Figure 5: $t = 1, dt = 10^{-3}, a = 1.5, \sigma = 0.5, h_1 = 0.7, h_2 = 0.5$ with $1,000$ substeps with 1st order asymptotic expansion. From top to bottom, the sample path, exact and approximated density functions, and the difference of the approximated densities from the exact one. In the middle graph, a red line labeled by "Benes" denotes the exact density function.

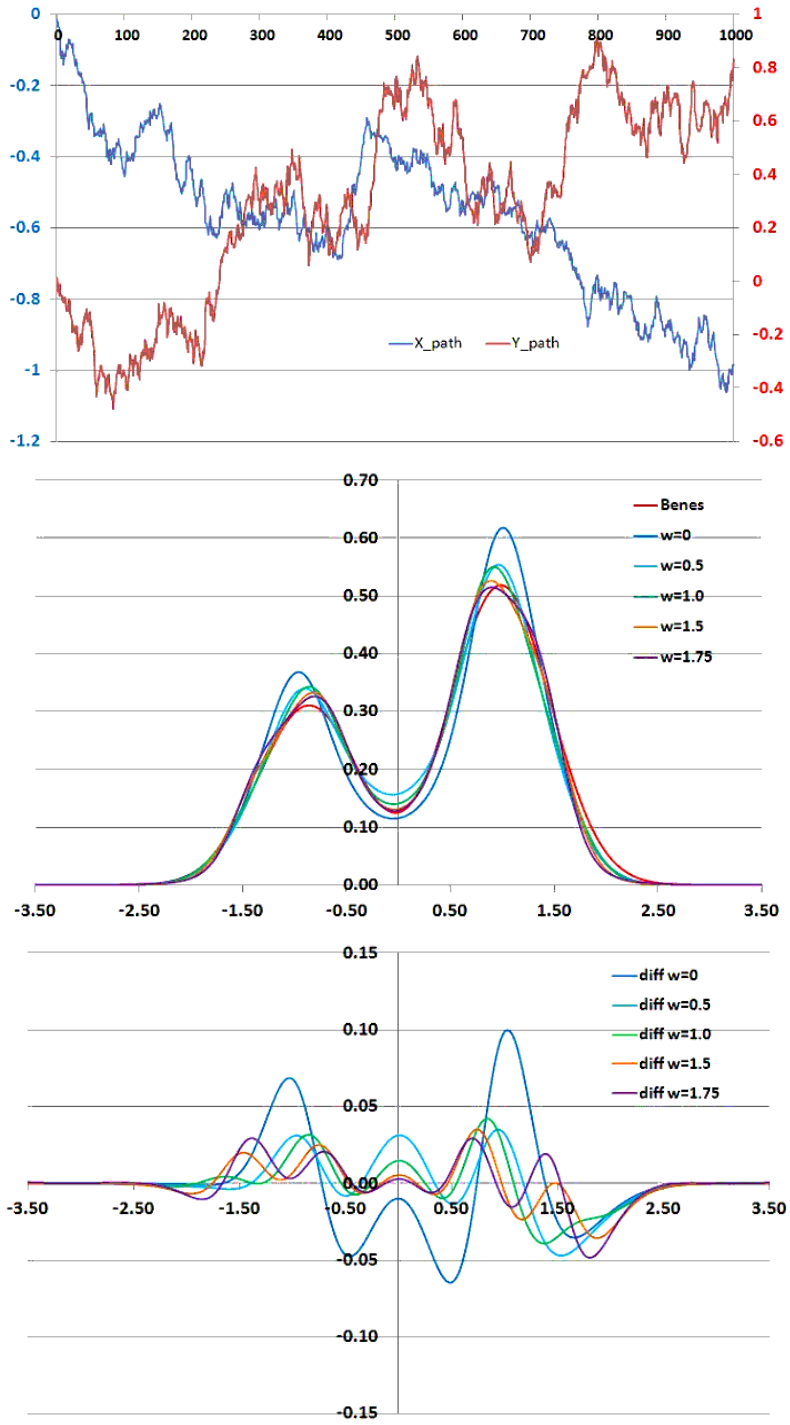

Figure 6: $t = 1, dt = 10^{-3}, a = 2.0, \sigma = 0.5, h_1 = 1.0, h_2 = 0.5$ with $1,000$ substeps with 1st order asymptotic expansion. From top to bottom, the sample path, exact and approximated density functions, and the difference of the approximated densities from the exact one. In the middle graph, a red line labeled by "Benes" denotes the exact density function.