Complete Real Time Solution of the General Nonlinear Filtering Problem without Memory

Xue Luo and Stephen S.-T. Yau, *Fellow, IEEE*

Dedicated to Professor Yu-Chi Ho and Professor Sanjoy Mitter on the occasion of their 80th birthday

Abstract—It is well known that the nonlinear filtering problem has important applications in both military and civil industries. The central problem of nonlinear filtering is to solve the Duncan-Mortensen-Zakai (DMZ) equation in real time and in a memoryless manner. In this paper, we shall extend the algorithm developed previously by S.-T. Yau and the second author to the most general setting of nonlinear filterings, where the explicit time-dependence is in the drift term, observation term, and the variance of the noises could be a matrix of functions of both time and the states. To preserve the off-line virture of the algorithm, necessary modifications are illustrated clearly. Moreover, it is shown rigorously that the approximated solution obtained by the algorithm converges to the real solution in the L^1 sense. And the precise error has been estimated. Finally, the numerical simulation support the feasibility and efficiency of our algorithm.

Index Terms—Nonlinear filtering, Duncan-Mortensen-Zakai equation, time-varying systems, convergence analysis.

I. INTRODUCTION

TRACING back to 1960s, two most influential mathematics papers [11], [12] have been published in ASME's Journal of Basic Engineering. These are so-called Kalman filter (KF) and Kalman-Bucy filter. They addressed a significant question: How does one get accurate estimate from noisy data? The applications of KF are endless, from seismology to bioengineering to econometrics. The KF surpasses the other filtering in, at least, the following two aspects:

- The KF uses each new obervation to update a probability distribution for the state of the system without refering back to any earlier observations. This is so-called "memoryless" or "without memory".
- The KF makes the decisions of the state on the spot, while the observation data keep coming in. This property is called "real time" application.

Despite its success in many real applications, the limitations on the nonlinearity and Gaussian assumption of the initial probability density of the KF push the mathematicians and scientists to seek the optimal nonlinear filtering. One direction is to modify KF to adapt the nonlinearities. The researchers developed extended Kalman filter (EKF), unscented Kalman

filter, ensemble Kalman filter, etc., which can handle weak nonlinearities (that is almost linear). But for serious nonlinearities, they may completely fail. The failure of EKF is shown in our numerical experiment, see the 1D cubic sensor in section VII.A.

Another direction, and also the most popular method nowadays, is the particle filter (PF), refer to as [1], [3] and reference therein. It is developed from sequential Monte Carlo method. On the one hand, the PF is appliable to nonlinear, non-Gaussian state update and observation equations. As the number of particles goes to infinity, the PF estimate becomes asymptotically optimal. On the other hand, it is hard to be implemented as a real time application, due to its essence of Monte Carlo simulation.

Besides the widely used two methods above, the partial differential equaitons (PDE) methods are introduced to the nonlinear filtering in 1960s. These methods are based on the fact that the unnormalized conditional density of the states is the solution of Duncan-Mortensen-Zakai (DMZ) equation, refer to as [6], [16] and [23]. The classical PDE methods could be applied to this stochastic PDE to obtain an approximation to the density. Yet, the main drawback of PDE methods are the intensive computation. It is almost impossible to achieve the "real time" performance. To overcome this shortcoming, the splitting-up algorithm is introduced to move the heavy computation off-line. It is like the Trotter product formula from semigroup theory. This operator splitting algorithm is proposed for the DMZ equation by Bensoussan, Glowinski, and Rascanu [5]. More research articles follow this direction are [9], [15] and [10], etc. Unfortunately, it is pointed out in [5] that the soundness of this algorithm is verified only to the filtering with bounded drift term and observation term (i.e., f and h in (2.1)). Essentially with the similar idea, Yau and Yau [22] developed a novel algorithm to the "pathwise-robust" DMZ equation (see (2.6)), where the boundedness conditions are weakened by some mild growth conditions on f and h. The two nice properties of the KF have also been kept in this algorithm: "without memory" and "real time". But their algorithm has only been rigorously proved in theory, when the drift term, the observation term (f, h in (2.1)) are not explicitly dependent on time, the variance of the noises (G in (2.1)) is the identity matrix, and the noises are standard Brownian motion processes ($S = I_{r \times r}$, $Q = I_{m \times m}$ in (2.1)).

In this paper, we shall extend the algorithm in [22] to the most general settings of nonlinear filtering problems, in the sense that the drift term, the observation term could explicitly depend on time, the variance of the noises S, Q are time-

dependent, and G could be a matrix of functions of both time and the states. We shall validate our algorithm under very mild growth conditions on f, h and G, see (3.5), (3.6) and (3.8). These are essentially time-dependent analogue of those in [22]. First of all, this extension is absolutely necessary. Many real applications have explicit time-dependence in their models, say the target orientation angles estimation from target position/velocity in constant turn model, where the angular velocities are piecewise constant functions in time [18]. Second, this extension is nontrivial from the mathematical point of view. More trickier analysis of PDE is required. For instance, we need to take care of the more general elliptic operator D_w^2, see (2.5), rather than the Laplacian.

This paper is organized in the following. The detailed formulation of our algorithm is described in section II; In section III, we state our main theorems which validate our algorithm in theory. Notations and prelimilary are in section IV. Section V is devoted to the proofs of the main theorems. The lower bound of the density function is investigated in section VI. Numerical simulations are included in section VII. Finally, we arrive the conclusion. The appendices is consisted of the proof of the well-posedness theorem and the proof of an interesting property of the density function.

II. MODEL AND ALGORITHM

The model we are considering is the signal observation model with explicit time-dependence in the drift term, observation term and the variance of the noises:

$$\begin{cases} dx_t = f(x_t,t)dt + G(x_t,t)dv_t, \\ dy_t = h(x_t,t)dt + dw_t, \end{cases} \quad (2.1)$$

where x_t and f are n-vectors, G is an $n \times r$ matrix, and v_t is an r-vector Brownian motion process with $E[dv_t dv_t^T] = Q(t)dt$, y_t and h are m-vectors and w_t is an m-vector Brownian motion process with $E[dw_t dw_t^T] = S(t)dt$ and $S(t) > 0$. We refer to x_t as the state of the system at time t with some initial state x_0 (not necessarily obeying Gaussian distribution) and y_t as the observation at time t with $y_0 = 0$. We assume that $\{v_t, t \geq 0\}$, $\{w_t, t \geq 0\}$ and x_0 are independent. For the sake of convenience, let us call this system is the "time-varying" case, while in [22] the "time-invariant" case is studied.

Throughout this paper, we assume that f, h and G are C^2 in space and C^1 in time. Some growth conditions on f and h are expected to guaratee the existence and uniqueness of the "pathwise-robust" DMZ equation.

The unnormalized density function $\sigma(x,t)$ of x_t conditioned on the observation history $Y_t = \{y_s : 0 \leq s \leq t\}$ satisfies the DMZ equation (for the detailed formulation, see [6])

$$\begin{cases} d\sigma(x,t) = L\sigma(x,t)dt + \sigma(x,t)h^T(x,t)S^{-1}(t)dy_t \\ \sigma(x,0) = \sigma_0(x), \end{cases} \quad (2.2)$$

where $\sigma_0(x)$ is the probability density of the initial state x_0, and

$$L(*) \equiv \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j}\left[(GQG^T)_{ij}*\right] - \sum_{i=1}^n \frac{\partial(f_i*)}{\partial x_i}. \quad (2.3)$$

In this paper, we don't solve the DMZ equation directly, due to the following two reasons. On the one hand, the DMZ equation (2.2) is a stochastic partial differential equation due to the term dy_t. There is no easy way to derive a recursive algorithm to solve this equation. On the other hand, in real applications, one may be more interested in constructing robust state estimators from each observation path, instead of having certain statistical data of thousands of repeated experiments. Here, the robustness means our state esitmator is not sensitive to the observation path. This property is important, since in most of the real applications, the observation arrives and is processed at discrete moments in time. The state estimator is expected to still perform well based on the linear interpolation of the discrete observations, instead of the real continuous observation path. For each "given" observation, making an invertible exponential transformation [19]

$$\sigma(x,t) = \exp[h^T(x,t)S^{-1}(t)y_t]\rho(x,t), \quad (2.4)$$

the DMZ equation is transformed into a deterministic partial differential equation (PDE) with stochastic coefficients, which we will refer as the "pathwise-robust" DMZ equation

$$\begin{cases} \dfrac{\partial\rho}{\partial t}(x,t) + \dfrac{\partial}{\partial t}(h^T S^{-1})^T y_t \rho(x,t) \\ \quad = \exp(-h^T S^{-1} y_t)\left[L - \dfrac{1}{2}h^T S^{-1}h\right] \\ \qquad \cdot [\exp(h^T S^{-1}y_t)\rho(x,t)] \\ \rho(x,0) = \sigma_0(x). \end{cases} \quad (2.5)$$

Or equivalently,

$$\begin{cases} \dfrac{\partial\rho}{\partial t}(x,t) = \dfrac{1}{2}D_w^2\rho(x,t) + F(x,t)\cdot\nabla\rho(x,t) + J(x,t)\rho(x,t) \\ \rho(x,0) = \sigma_0(x), \end{cases}$$
$$(2.6)$$

where

$$D_w^2 = \sum_{i,j=1}^n (GQG^T)_{ij}\frac{\partial^2}{\partial x_i \partial x_j}, \quad (2.7)$$

$$F(x,t) = \left[\sum_{j=1}^n \frac{\partial}{\partial x_j}(GQG^T)_{ij} \right.$$
$$\left. + \sum_{j=1}^n (GQG^T)_{ij}\frac{\partial K}{\partial x_j} - f_i\right]_{i=1}^n, \quad (2.8)$$

$$J(x,t) = -\frac{\partial}{\partial t}(h^T S^{-1})^T y_t + \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j}(GQG^T)_{ij}$$
$$+ \sum_{i,j=1}^n \frac{\partial}{\partial x_i}(GQG^T)_{ij}\frac{\partial K}{\partial x_j}$$
$$+ \frac{1}{2}\sum_{i,j=1}^n (GQG^T)_{ij}\left[\frac{\partial^2 K}{\partial x_i \partial x_j} + \frac{\partial K}{\partial x_i}\frac{\partial K}{\partial x_j}\right]$$
$$- \sum_{i=1}^n \frac{\partial f_i}{\partial x_i} - \sum_{i=1}^n f_i\frac{\partial K}{\partial x_i} - \frac{1}{2}(h^T S^{-1}h), \quad (2.9)$$

in which

$$K(x,t) = h^T(x,t)S^{-1}(t)y_t. \quad (2.10)$$

The existence and uniqueness of the "pathwise-robust" DMZ equation (2.6) has been investigated by Pardoux [17], Fleming-Mitter [7], Baras-Blankenship-Hopkins [2] and Yau-Yau [21], [22]. The well-posedness is guaranteed, when the drift term $f \in C^1$ and the observation term $h \in C^2$ are bounded in [17]. Fleming and Mitter treated the case where f and ∇f are bounded. Baras, Blankenship and Hopkins obtained the well-posedness result on the "pathwise-robust" DMZ equation with a class of unbounded coefficients only in one dimension. In [21], Yau and Yau established the well-posedness result under the condition that f, h have at most linear growth. In the appendices of [22], Yau and Yau obtained the existence and uniqueness results in the weighted Sobolev space, where f and h satisfy some mild growth condition. It is necessary to point out that there is a gap in their proof of existence (Theorem A.4). In this paper, we circumvent the gap by more delicate analysis to give a time-dependent analogous well-posedness result to the "pathwise-robust" DMZ equation under some mild growth conditions on f and h in Theorem 4.3.

The exact solution to (2.5) or (2.6), generally speaking, doesn't have a closed form. So many mathematicians pay their effort on seeking an efficient algorithm to construct a good approximation. In this paper, we will extend the algorithm in [22] to the "time-varying" case (cf. (2.1)). We will not only give the theoretical proof of the soundness of our algorithm, but also illustrate a "time-varying" numerical simulation to support our results. The difficulties are two folds: on one hand, the well-posedness of the "time-invariant" robust DMZ equation, under some conditions, has been investigated by [7], [17], [22], etc., while that in the "time-varying" case hasn't been established yet; on the other hand, the "time-varying" case will lead to more involved computations and more delicate analysis. For instance, the Laplacian in "time-invariant" case is replaced by a time-dependent elliptic operator D_w^2 in (2.7). Furthermore, the two nice properties of KF, namely "memoryless" and "real time", are preserved in our algorithm.

Let us assume that we know the observation time sequence $0 = \tau_0 < \tau_1 < \cdots < \tau_k = T$ apriorily. But the observation data $\{y_{\tau_i}\}$ at each sampling time τ_i, $i = 0, \cdots, k$ are unknown until the on-line experiment runs. We call the computation "off-line", if it can be performed without any on-line experimental data (or say pre-computed); otherwise, it is called "on-line" computations. One only concerns the computational complexity of the on-line computations, since this hinges the success of "real time" application.

Let us denote the observation time sequence as $\mathcal{P}_k = \{0 = \tau_0 < \tau_1 < \cdots < \tau_k = T\}$. Let ρ_i be the solution of the robust DMZ equation with $y_t = y_{\tau_{i-1}}$ on the interval $\tau_{i-1} \leq t \leq \tau_i$,

$i = 1, 2, \cdots, k$

$$\begin{cases} \dfrac{\partial \rho_i}{\partial t}(x,t) + \dfrac{\partial}{\partial t}\left(h^T S^{-1}\right)^T y_{\tau_{i-1}} \rho_i(x,t) \\ \quad = \exp\left(-h^T S^{-1} y_{\tau_{i-1}}\right)\left[L - \dfrac{1}{2}h^T S^{-1} h\right] \\ \qquad \cdot \left[\exp\left(h^T S^{-1} y_{\tau_{i-1}}\right)\rho_i(x,t)\right] \\ \rho_1(x,0) = \sigma_0(x), \\ \text{or} \\ \rho_i(x,\tau_{i-1}) = \rho_{i-1}(x,\tau_{i-1}), \quad \text{for } i = 2, 3, \cdots, k. \end{cases}$$
$$(2.11)$$

Define the norm of $\mathcal{P}_k$ by $|\mathcal{P}_k| = \sup_{1 \leq i \leq k}(\tau_i - \tau_{i-1})$. Intuitively, as $|\mathcal{P}_k| \to 0$, we have

$$\sum_{i=1}^{k} \chi_{[\tau_{i-1}, \tau_i]}(t)\rho_i(x,t) \to \rho(x,t)$$

in some sense, for all $0 \leq t \leq T$, where $\rho(x,t)$ is the exact solution of (2.5). That is to say, intuitively, the denser the sampling time sequence is, the more accurate the approximate solution should be obtained. Even though the intuition is shown rigorously to be true, it is impractical to solve (2.11) in the "real time" manner, since the "on-line" data $\{y_{\tau_i}\}$, $i = 1, \cdots, k$, are contained in the coefficients of (2.11). Therefore, we have to numerically solve the time-consuming PDE on-line, every time after the new observation data coming in. Yet, the proposition below helps to move the heavy computations off-line. This is the key ingredient of the algorithm in [22], so is in ours.

Proposition 2.1: For each $\tau_{i-1} \leq t < \tau_i$, $i = 1, 2, \cdots, k$, $\rho_i(x,t)$ satisfies (2.11) if and only if

$$u_i(x,t) = \exp\left[h^T(x,t)S^{-1}(t)y_{\tau_{i-1}}\right]\rho_i(x,t), \quad (2.12)$$

satisfies the Kolmogorov forward equation (KFE)

$$\frac{\partial u_i}{\partial t}(x,t) = \left(L - \frac{1}{2}h^T S^{-1} h\right)u_i(x,t), \quad (2.13)$$

where L is defined in (2.3).

It is clear that (2.13) is independent of the observation path $\{y_{\tau_i}\}_{i=0}^{k}$, and the transformation between u_i and ρ_i is one-to-one. It is also not hard to see that (2.13) could be numerically solved beforehand. Observe that the operator $\left(L - \frac{1}{2}h^T S^{-1} h\right)$ is time-varying, unlike that studied in [22]. Let us denote it as $U(t)$ for short and emphasis its time-dependence. But this doesn't affect the "off-line" virture of our algorithm. Under certain conditions, $\{U(t)\}_{t \in [0,T]}$ forms a family of strong elliptic operators. Furthermore, the operator $U(t) : D(U(t)) \subset L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is the infinitesimal generator of the two-parameter semigroup $\mathcal{U}(t,\tau)$, for $t \geq \tau$. In particular, with the observation time sequence known $\{\tau_i\}_{i=1}^{k}$, we obtain a sequence of two-parameter semigroup $\{\mathcal{U}(t,\tau_{i-1})\}_{i=1}^{k}$, for $\tau_{i-1} \leq t < \tau_i$. Let us take the initial conditions of KFE (2.13) at $t = \tau_i$ as a set of complete orthonormal base in $L^2(\mathbb{R}^n)$, say $\{\phi_l(x)\}_{l=1}^{\infty}$. We pre-compute the solutions of (2.13) at time $t = \tau_{i+1}$, denoted as $\{\mathcal{U}(\tau_{i+1},\tau_i)\phi_l\}_{l=1}^{\infty}$. These data should be stored in preparation of the on-line computations. Compared with the "time-invariant" case, the price to pay is that the "time-varying" case requires more storage capacity,

since $\{\mathcal{U}(\tau_{i+1}, \tau_i)\phi_l\}_{l=1}^{\infty}$ differs from each τ_i, $i = 1, \cdots, k$, and all of them need to be stored. In general, the longer simulation time is the more storage it requires in the "time-varying" case. While the storage of the data is independent of the simulation time in the "time-invariant" case. Nevertheless, it won't affect the off-line virture of our algorithm.

The on-line computation in our algorithm is consisted of two parts at each time step τ_{i-1}, $i = 1, \cdots, k$.

- Project the initial condition $u_i(x, \tau_{i-1}) \in L^2(\mathbb{R}^n)$ at $t = \tau_{i-1}$ onto the base $\{\phi_l(x)\}_{l=1}^{\infty}$, i.e., $u_i(x, \tau_{i-1}) = \sum_{l=1}^{\infty} \hat{u}_{i,l}\phi_l(x)$. Hence, the solution to (2.13) at $t = \tau_i$ can be expressed as

$$
\begin{aligned}
u_i(x, \tau_i) &= \mathcal{U}(\tau_i, \tau_{i-1})u_i(x, \tau_{i-1}) \\
&= \sum_{l=1}^{\infty} \hat{u}_{i,l}\left[\mathcal{U}(\tau_i, \tau_{i-1})\phi_l(x)\right], \quad (2.14)
\end{aligned}
$$

where $\{\mathcal{U}(\tau_i, \tau_{i-1})\phi_l(x)\}_{l=1}^{\infty}$ have already been computed off-line.

- Update the initial condition of (2.13) at τ_i with the new observation y_{τ_i}. Let us specify the observation updates (the initial condition of (2.13)) for each time step. For $0 \le t \le \tau_1$, the initial condition is $u_1(x, 0) = \sigma_0(x)$. At time $t = \tau_1$, when the observation y_{τ_1} is available,

$$
\begin{aligned}
u_2(x, \tau_1) &\overset{(2.12)}{=} \exp\left[h^T(x, \tau_1)S^{-1}(\tau_1)y_{\tau_1}\right]\rho_2(x, \tau_1) \\
&\overset{(2.12),(2.11)}{=} \exp\left[h^T(x, \tau_1)S^{-1}(\tau_1)y_{\tau_1}\right]u_1(x, \tau_1),
\end{aligned}
$$

with the fact $y_0 = 0$. Here, $u_1(x, \tau_1) = \sum_{l=1}^{\infty} \hat{u}_{1,l}\left[\mathcal{U}(\tau_1, 0)\phi_l(x)\right]$, where $\{\hat{u}_{1,l}\}_{l=1}^{\infty}$ is computed in the previous step, and $\{\mathcal{U}(\tau_1, 0)\phi_l(x)\}_{l=1}^{\infty}$ are prepared by off-line computations. Hence, we obtain the initial condition $u_2(x, \tau_1)$ of (2.13) for the next time interval $\tau_1 \le t \le \tau_2$. Recursively, the initial condition of (2.13) for $\tau_{i-1} \le t \le \tau_i$ is

$$
\begin{aligned}
u_i(x, \tau_{i-1}) = &\exp\left[h^T(x, \tau_{i-1})S^{-1}(\tau_{i-1})(y_{\tau_{i-1}} - y_{\tau_{i-2}})\right] \\
&\cdot u_{i-1}(x, \tau_{i-1}), \quad (2.15)
\end{aligned}
$$

for $i = 2, 3, \cdots, k$, where $u_{i-1}(x, \tau_{i-1}) = \sum_{l=1}^{\infty} \hat{u}_{i-2,l}\left[\mathcal{U}(\tau_{i-1}, \tau_{i-2})\phi_l(x)\right]$.

The approximation of $\rho(x, t)$, denoted as $\hat{\rho}(x, t)$, is obtained

$$
\hat{\rho}(x, t) = \sum_{i=1}^{k} \chi_{[\tau_{i-1}, \tau_i]}(t)\rho_i(x, t), \quad (2.16)
$$

where $\rho_i(x, t)$ is obtained from $u_i(x, t)$ by (2.12). And $\sigma(x, t)$ could be recovered by (2.4).

A natural question comes to us:

Is $\hat{\rho}(x, t)$, obtained by our algorithm, a good approximation of the exact solution $\rho(x, t)$ to (2.5), for $(x, t) \in \mathbb{R}^n \times [0, T]$, as $|\mathcal{P}_k| \to 0$? If it is, then in what sense?

III. STATEMENTS OF THE MAIN THEOREMS

In this section, we shall state the main theorems in this paper, which validate our algorithm in theory. Notice that u_i in (2.13) and ρ_i in (2.11) are one-to-one. Hence, we shall deal with ρ_i in the sequel.

We first show that the exact solution ρ of the "pathwise-robust" DMZ equation (2.5) is well approximated by ρ_R as $R \to \infty$, for any $t \in [0, T]$, where ρ_R is the solution to (2.5) restricted on B_R (the ball centered at the origin with the radius R):

$$
\begin{cases}
\dfrac{\partial \rho_R}{\partial t}(x, t) = \dfrac{1}{2}D_w^2\rho_R(x, t) + F(x, t)\nabla\rho_R(x, t) \\
\qquad\qquad + J(x, t)\rho_R(x, t) \\
\rho_R(x, 0) = \sigma_{0, B_R}(x) \\
\rho_R(x, t) = 0 \qquad \text{for } (x, t) \in \partial B_R \times [0, T],
\end{cases}
\quad (3.1)
$$

where D_w^2, $F(x, t)$ and $J(x, t)$ are defined in (2.7)-(2.9) and $\sigma_{0, \Omega}$ is defined as

$$
\sigma_{0, \Omega}(x) = \begin{cases}
\sigma_0(x), & x \in \Omega_\epsilon \\
\text{smooth}, & x \in \Omega \setminus \Omega_\epsilon \\
0, & x \in \mathbb{R}^n \setminus \Omega,
\end{cases}
\quad (3.2)
$$

in which $\Omega_\epsilon = \{x \in \Omega : \text{dist}(x, \partial\Omega) > \epsilon\}$. Next, it is left to show that ρ_R is well approximated by the solution obtained by our algorithm restricted on B_R. In fact, on the time interval $[0, \tau]$, $0 < \tau \le T$. Let us denote the time partition $\mathcal{P}_k^\tau = \{0 = \tau_0 < \tau_1 < \cdots < \tau_k = \tau\}$. $\rho_R(x, \tau)$ is well approximated by $\rho_{k,R}(x, \tau)$, as $k \to +\infty$, in the L^1 sense, where $\rho_{k,R}$ is the solution of (2.11) restricted on B_R.

For the notational convenience, let us denote

$$
\begin{aligned}
N(x, t) \equiv &-\dfrac{\partial}{\partial t}\left(h^T S^{-1}\right)y_t - \dfrac{1}{2}D_w^2 K \\
&+ \dfrac{1}{2}D_w K \cdot \nabla K - f \cdot \nabla K - \dfrac{1}{2}\left(h^T S^{-1}h\right), \quad (3.3)
\end{aligned}
$$

where

$$
D_w * = \left[\sum_{j=1}^{n}\left(GQG^T\right)_{ij}(x, t)\dfrac{\partial *}{\partial x_j}\right]_{i=1}^{n}, \quad (3.4)
$$

and D_w^2 and K are defined in (2.7) and (2.10), respectively.

The error estimate between ρ and ρ_R is given by the following theorem.

Theorem 3.1: For any $T > 0$, let $\rho(x, t)$ be a solution of the "pathwise-robust" DMZ equation (2.6) in $\mathbb{R}^n \times [0, T]$. Let $R \gg 1$ and ρ_R be the solution to (3.1). Assume the following conditions are satisfied, for all $(x, t) \in \mathbb{R}^n \times [0, T]$:

1) $N(x, t) + \dfrac{3}{2}n\left|\left|GQG^T\right|\right|_\infty + |f - D_w K| \le C$, $\quad (3.5)$
2) $e^{-\sqrt{1+|x|^2}}\left[14n\left|\left|GQG^T\right|\right|_\infty + 4|f - D_w K|\right] \le \tilde{C}$, $\quad (3.6)$

where N, D_w and K are defined in (3.3), (3.4) and (2.10), respectively, and C, $\tilde{C}$ are constants possibly depending on T. Let $v = \rho - \rho_R$, then $v \ge 0$ for all $(x, t) \in B_R \times [0, T]$ and

$$
\int_{B_{\frac{R}{2}}} v(x, T) \le \bar{C}e^{-\frac{9}{16}R}\int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}}\sigma_0(x), \quad (3.7)
$$

where $\bar{C}$ is some constant, which may depend on T.

The next theorem tells us that ρ_R is well approximated by the solution obtained by our algorithm restricted on B_R. More generally, B_R is replaced by any bounded domain $\Omega \subset \mathbb{R}^n$ in the theorem.

Theorem 3.2: Let Ω be a bounded domain in $\mathbb{R}^n$. Assume that

1) $|N(x,t)| \leq C$, $\qquad\qquad\qquad\qquad\qquad\quad$ (3.8)
2) There exists some $\alpha \in (0,1)$, such that

$$|N(x,t) - N(x,t;\bar{t})| \leq \tilde{C}|t - \bar{t}|^{\alpha}, \qquad (3.9)$$

for all $(x,t) \in \Omega \times [0,T]$, $\bar{t} \in [0,T]$, where $N(x,t)$ is in (3.3), and $N(x,t;\bar{t})$ denotes $N(x,t)$ with the observation $y_t = y_{\bar{t}}$. Let $\rho_\Omega(x,t)$ be the solution of (2.6) on $\Omega \times [0,T]$ with $0-$Dirichlet boundary condition:

$$\begin{cases} \dfrac{\partial \rho_\Omega}{\partial t}(x,t) = \dfrac{1}{2}D_w^2\rho_\Omega(x,t) + F(x,t)\cdot\nabla\rho_\Omega(x,t) \\ \qquad\qquad + J(x,t)\rho_\Omega(x,t) \\ \rho_\Omega(x,0) = \sigma_{0,\Omega}(x) \\ \rho_\Omega(x,t)|_{\partial\Omega} = 0, \end{cases} \quad (3.10)$$

where D_w^2, $F(x,t)$ and $J(x,t)$ are defined in (2.7)-(2.9) and $\sigma_{0,\Omega}$ is defined in (3.2). For any $0 \leq \tau \leq T$, let $\mathcal{P}_k^\tau = \{0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_k = \tau\}$ be a partition of $[0,\tau]$, where $\tau_i = \frac{i\tau}{k}$. Let $\rho_{i,\Omega}(x,t)$ be the approximate solution obtained by our algorithm restricted on $\Omega \times [\tau_{i-1}, \tau_i]$. Or equivalently, $\rho_{i,\Omega}$ is the solution on $\Omega \times [\tau_{i-1}, \tau_i]$ of the equation

$$\begin{cases} \dfrac{\partial \rho_{i,\Omega}}{\partial t}(x,t) = \dfrac{1}{2}D_w^2\rho_{i,\Omega}(x,t) + F(x,t;\tau_{i-1})\cdot\nabla\rho_{i,\Omega}(x,t) \\ \qquad\qquad + J(x,t;\tau_{i-1})\rho_{i,\Omega}(x,t) \\ \rho_{i,\Omega}(x,\tau_{i-1}) = \rho_{i-1,\Omega}(x,\tau_{i-1}) \\ \rho_{i,\Omega}(x,t)|_{\partial\Omega} = 0, \end{cases}$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (3.11)$$

for $i = 1, 2, \cdots, k$, with the convention that $\rho_{1,\Omega}(x,0) = \sigma_{0,\Omega}(x)$. Here, $F(x,t;\tau_{i-1})$, $J(x,t;\tau_{i-1})$ denote $F(x,t)$, $J(x,t)$ with the observation $y_t = y_{\tau_{i-1}}$, respectively. Then

$$\rho_\Omega(x,\tau) = \lim_{k\to\infty} \rho_{k,\Omega}(x,\tau),$$

in the L^1 sense in space and the following estimate holds:

$$\int_\Omega |\rho_\Omega - \rho_{k,\Omega}|(x,\tau) \leq \bar{C}\frac{1}{k^\alpha}, \qquad (3.12)$$

where $\bar{C}$ is a generic constant, depending on T, $\int_\Omega \sigma_{0,\Omega}$. The right-hand side of (3.12) tends to zero as $k \to \infty$.

IV. NOTATIONS AND PRELIMINARY

Throughout the paper, let $\mathbb{Q}_T = \mathbb{R}^n \times [0,T]$. Let $H^1(\mathbb{R}^n)$ be the Sobolev space, equipped the norm

$$||u(x)||_1^2 = \int_{\mathbb{R}^n}(u^2 + |\nabla_x u|^2)dx.$$

And let $H^{1;1}(\mathbb{Q}_T)$ be the functional space of both t and x, with the norm

$$||v(x,t)||_{1;1}^2 = \int_{\mathbb{Q}_T}(v^2 + |\nabla_x v|^2 + |\partial_t v|^2)dxdt.$$

The subspace of $H^{1;1}(\mathbb{Q}_T)$ consisting of functions $v(x,t)$ which have compact supports in $\mathbb{R}^n$ for any t is denoted as $H_0^{1;1}(\mathbb{Q}_T)$.

Definition 4.1: The function $u(x,t)$ in $H_0^{1;1}(\mathbb{Q}_T)$ is called a *weak solution* of the initial value problem

$$\begin{cases} \displaystyle\sum_{i,j=1}^n \dfrac{\partial}{\partial x_i}\left(A_{ij}(x,t)\dfrac{\partial u}{\partial x_j}\right) + \sum_{i=1}^n B_i(x,t)\dfrac{\partial u}{\partial x_i} \\ \qquad + C(x,t)u = \dfrac{\partial u}{\partial t}, \\ u(x,0) = u_0(x) \end{cases}$$

if for any function $\Phi(x,t) \in H_0^{1;1}(\mathbb{Q}_T)$ the following relation holds:

$$\int\int_{\mathbb{Q}_T}\left[\sum_{i,j=1}^n A_{ij}\dfrac{\partial u}{\partial x_i}\dfrac{\partial \Phi}{\partial x_j} - \left(\sum_{i=1}^n B_i\dfrac{\partial u}{\partial x_i} + Cu + \dfrac{\partial u}{\partial t}\right)\Phi\right]dxdt = 0$$

and $u(x,0) = u_0(x)$.

We assume that the following conditions hold throughout the paper:

1) The operator L defined in (2.3) is a strong elliptic operator and it is bounded from above on $\mathbb{Q}_T$. That is, there exists a constant $\lambda > 0$ such that

$$\lambda|\xi|^2 \leq \sum_{i,j=1}^n (GQG^T)_{ij}\xi_i\xi_j,$$

for any $(x,t) \in \mathbb{Q}_T$, for any $\xi = (\xi_1, \xi_2, \cdots, \xi_n) \in \mathbb{R}^n$. And

$$||GQG^T||_\infty = \sup_{(x,t)\in\mathbb{Q}_T}|GQG^T|_\infty < \infty,$$

where $|\cdot|_\infty$ is the sup-norm of the matrix.
2) The initial density function $\sigma_0(x) \in H^1(\mathbb{R}^n)$ decays fast enough. To be more specific, we require that

$$\int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}}\sigma_0(x) < \infty.$$

A. Existence and uniqueness of the non-negative weak solution

For the sake of completeness, we state the existence and uniqueness of the non-negative weak solution to (2.6) on $\mathbb{Q}_T$ below.

Theorem 4.3: Under the conditions 1)-2) and the conditions (A.1)-(A.5) in Theorem A.1, the "pathwise-robust" DMZ equation (2.6) on $\mathbb{Q}_T$ with the initial value $\sigma_0 \in H^1(\mathbb{R}^n)$ admits a non-negative weak solution $\rho \in H^{1;1}(\mathbb{Q}_T)$. Assume further that the conditions (A.21)-(A.23) in Theorem A.2 are satisfied, then the weak solution ρ on $\mathbb{Q}_T$ is unique.

To avoid the distraction from our main theorems, we leave the detailed proof of this theorem in Appendix A.

B. Technical lemma

In the proofs of our main theorems, we will repeatedly adopt the following lemma with suitably chosen test functions. Let us state the lemma and sketch the proof here.

Lemma 4.1: Assume that ρ_Ω satisfies the "pathwise-robust" DMZ equation (2.6) on some bounded domain $\Omega \in \mathbb{R}^n$, for $0 \le t \le T$. Then, for any test function $\psi(x) \in C^\infty(\Omega)$, we have

$$
\begin{aligned}
\frac{d}{dt} \int_\Omega \psi \rho_\Omega =& \frac{1}{2} \int_\Omega D_w^2 \psi \rho_\Omega + \int_\Omega (f - D_w K) \cdot \nabla \psi \rho_\Omega \\
&+ \int_\Omega \psi \rho_\Omega N + \frac{1}{2} \int_{\partial\Omega} \psi \left(D_w \rho_\Omega \cdot \nu \right) \\
&- \frac{1}{2} \int_{\partial\Omega} \rho_\Omega \left(D_w \psi \cdot \nu \right) \\
&+ \frac{1}{2} \int_{\partial\Omega} \psi \rho_\Omega \sum_{i,j=1}^n \frac{\partial}{\partial x_i} \left(GQG^T \right)_{ij} \nu_j \\
&+ \int_{\partial\Omega} \psi \rho_\Omega \left(D_w K \cdot \nu \right) - \int_{\partial\Omega} \psi \rho_\Omega (f \cdot \nu),
\end{aligned}
\tag{4.1}
$$

where $\nu = (\nu_1, \nu_2, \cdots, \nu_n)$ is the exterior normal vector of Ω, and D_w^2, K and D_w are defined in (2.7), (2.10) and (3.4), respectively.

Sketch of the proof: Multiply $\psi(x)$ on both sides of (2.6) and integrate over the domain Ω, it yields

$$
\frac{d}{dt} \int_\Omega \psi \rho_\Omega = \int_\Omega \psi \left[\frac{1}{2} D_w^2 \rho_\Omega + F(x,t) \cdot \nabla \rho_\Omega + J(x,t) \rho_\Omega \right],
\tag{4.2}
$$

where $F(x,t)$ and $J(x,t)$ are defined in (2.8) and (2.9), respectively. After applying integration by parts to the first two terms on the right-hand side of (4.2), (4.1) is obtained by written in compact notations. ∎

V. PROOFS OF THE MAIN THEOREMS

A. Reduction to the bounded domain case

In this section we shall prove that the solution ρ to the "pathwise-robust" DMZ equation (2.6) in $\mathbb{R}^n$ can be well approximated by the solution ρ_R of (3.1) in a large ball B_R. Moreover, the error estimate with respect to the radius R is given explicitly in the L^1 sense. Let C, $\tilde{C}$ and $\hat{C}$ denote the generic constants, which may differ from line to line.

We first show an interesting proposition, which reflects how the density function in the large ball changing with respect to time. It is also an important ingredient of the error estimate in Theorem 3.1.

Proposition 5.2: For any $T > 0$, let $\rho_R(x,t)$ be a solution of the "pathwise-robust" DMZ equation restricted on B_R (3.1). Assume that condition (3.5) is satisfied. Then

$$
\int_{B_R} e^{\sqrt{1+|x|^2}} \rho_R(x,t) \le e^{Ct} \int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}} \sigma_0(x).
\tag{5.1}
$$

Proof: Choose the test function in Lemma 4.1 $\psi = e^{\phi_1}$, where $\phi_1 \in C^\infty(B_R)$, $B_R = \{x \in \mathbb{R}^n : |x| \le R\}$. Let ρ_R be the solution of the "pathwise-robust" DMZ equation (3.1) on the ball B_R. By Lemma 4.1, we have

$$
\begin{aligned}
\frac{d}{dt} \int_{B_R} e^{\phi_1} \rho_R =& \int_{B_R} e^{\phi_1} \rho_R \left[\frac{1}{2} \left(D_w^2 \phi_1 + D_w \phi_1 \cdot \nabla \phi_1 \right) \right. \\
&\left. + (f - D_w K) \cdot \nabla \phi_1 + N \right] \\
&+ \frac{1}{2} \int_{\partial B_R} e^{\phi_1} \left(D_w \rho_R \cdot \nu \right).
\end{aligned}
\tag{5.2}
$$

All the boundary integrals in (4.1) vanish, except the first term in (4.1), since $\rho_R|_{\partial\Omega} = 0$. Moreover, recall that $\rho_R \ge 0$ in B_R and vanishes on ∂B_R implies that $\frac{\partial \rho_R}{\partial \nu}|_{\partial B_R} \le 0$. Hence, on ∂B_R,

$$
\begin{aligned}
(D_w \rho_R \cdot \nu) &= \sum_{i=1}^n \left[\sum_{j=1}^n (GQG^T)_{ij} \frac{\partial \rho_R}{\partial r} \frac{\partial r}{\partial x_j} \right] \nu_i \\
&= \frac{\partial \rho_R}{\partial r} \left[\sum_{i,j=1}^n (GQG^T)_{ij} \frac{x_j}{r} \frac{x_i}{r} \right] \le 0,
\end{aligned}
$$

by the positive definite assumption of (GQG^T). Thus, (5.2) can be reduced further

$$
\begin{aligned}
\frac{d}{dt} \int_{B_R} e^{\phi_1} \rho_R \le & \int_{B_R} e^{\phi_1} \rho_R \left[\frac{1}{2} \left(D_w^2 \phi_1 + D_w \phi_1 \cdot \nabla \phi_1 \right) \right. \\
&\left. + (f - D_w K) \cdot \nabla \phi_1 + N \right].
\end{aligned}
\tag{5.3}
$$

Choose $\phi_1(x) = \sqrt{1+|x|^2}$ and estimate the terms containing ϕ_1 on the right-hand side of (5.3) one by one:

$$
\begin{aligned}
D_w^2 \phi_1 =& \sum_{i=1}^n \left(GQG^T \right)_{ii} \frac{1}{\sqrt{1+|x|^2}} \\
&- \sum_{i,j=1}^n \left(GQG^T \right)_{ij} \frac{x_i x_j}{(1+|x|^2)^{\frac{3}{2}}} \\
\le & \left\| GQG^T \right\|_\infty \left[\frac{n}{\sqrt{1+|x|^2}} + \frac{n|x|^2}{(1+|x|^2)^{\frac{3}{2}}} \right] \\
\le & 2n \left\| GQG^T \right\|_\infty,
\end{aligned}
\tag{5.4}
$$

$$
\begin{aligned}
D_w \phi_1 \cdot \nabla \phi_1 =& \sum_{i,j=1}^n \left(GQG^T \right)_{ij} \frac{x_i x_j}{1+|x|^2} \\
\le & \left\| GQG^T \right\|_\infty \frac{\sum_{i,j=1}^n x_i x_j}{1+|x|^2} \le n \left\| GQG^T \right\|_\infty,
\end{aligned}
\tag{5.5}
$$

and

$$
\begin{aligned}
\left| (f - D_w K) \cdot \nabla \phi_1 \right| \le & |f - D_w K| \cdot \frac{|x|}{\sqrt{1+|x|^2}} \\
\le & |f - D_w K|,
\end{aligned}
\tag{5.6}
$$

where $|\cdot|$ is the Euclidean norm. Substitute the estimate (5.4)-(5.6) back into (5.3), we get

$$
\begin{aligned}
& \frac{d}{dt} \int_{B_R} e^{\phi_1} \rho_R \\
\le & \int_{B_R} e^{\phi_1} \rho_R \left[\frac{3}{2} n \left\| GQG^T \right\|_\infty + |f - D_w K| + N \right] \\
\le & C \int_{B_R} e^{\phi_1} \rho_R,
\end{aligned}
$$

by condition (3.5). Hence,

$$
\begin{aligned}
\int_{B_R} e^{\phi_1} \rho_R(x,t) \le & e^{Ct} \int_{B_R} e^{\phi_1} \rho_R(x,0) \le e^{Ct} \int_{\mathbb{R}^n} e^{\phi_1} \rho(x,0) \\
= & e^{Ct} \int_{\mathbb{R}^n} e^{\phi_1} \sigma_0(x),
\end{aligned}
$$

for $0 \le t \le T$. ∎

We are ready to show Theorem 3.1, i.e. the solution ρ_R to (3.1) on B_R is a good approximation of ρ, the solution to (2.6) in $\mathbb{R}^n$.

Proof of Theorem 3.1: By the maximum principle (cf. Theorem 1, [8]), we have $v = \rho - \rho_R \geq 0$ for $(x, t) \in B_R \times [0, T]$, since $v|_{\partial B_R} \geq 0$ for $0 \leq t \leq T$. Let us choose ψ in Lemma 4.1 as

$$\varrho(x) = e^{-\phi_2(x)} - e^{-R},$$

where ϕ_2 is a radial symmetric function such that $\phi_2(x)|_{\partial B_R} = R$, $\nabla \phi_2|_{\partial B_R} = 0$ and ϕ_2 is increasing in $|x|$. Hence, $\varrho|_{\partial B_R} = 0$ and $\nabla \varrho|_{\partial B_R} = 0$. Apply Lemma 4.1 to v, taking the place of ρ_Ω, with the test function $\psi = \varrho$, we have

$$\frac{d}{dt} \int_{B_R} \varrho v = \int_{B_R} v \left[\frac{1}{2} D_w^2 \varrho + (f - D_w K) \cdot \nabla \varrho + \varrho N \right]$$

$$= \int_{B_R} v \left\{ \frac{1}{2} e^{-\phi_2} \left(D_w \phi_2 \cdot \nabla \phi_2 - D_w^2 \phi_2 \right) \right.$$

$$\left. - e^{-\phi_2} (f - D_w K) \cdot \nabla \phi_2 + \varrho N \right\}$$

$$= \int_{B_R} v \varrho \left[-\frac{1}{2} D_w^2 \phi_2 + \frac{1}{2} D_w \phi_2 \cdot \nabla \phi_2 \right.$$

$$\left. - (f - D_w K) \cdot \nabla \phi_2 + N \right]$$

$$+ e^{-R} \int_{B_R} e^{\sqrt{1+|x|^2}} v \left[e^{-\sqrt{1+|x|^2}} \right.$$

$$\cdot \left(-\frac{1}{2} D_w^2 \phi_2 + \frac{1}{2} D_w \phi_2 \cdot \nabla \phi_2 \right.$$

$$\left. - (f - D_w K) \cdot \nabla \phi_2) \right]$$

$$\triangleq \int_{B_R} v \varrho \mathrm{I}_1 + e^{-R} \int_{B_R} e^{\sqrt{1+|x|^2}} v \mathrm{I}_2.$$

Let us choose $\phi_2(x)$ in $\varrho(x)$ to be $\phi_2(x) = R\vartheta(\frac{|x|^2}{R^2})$, where $\vartheta(x) = 1 - (1-x)^2$. It is easy to check that $\phi_2(x)$ satisfies all the conditions we mentioned before. Direct computations yield, for any $x \in B_R$, $R >> 1$,

$$\left| D_w^2 \phi_2 \right| = \left| \sum_{i,j=1}^n \left(GQG^T \right)_{ij} \left(-\frac{8x_i x_j}{R^3} \right) \right.$$

$$\left. + \sum_{i=1}^n \left(GQG^T \right)_{ii} \frac{4}{R} \left(1 - \frac{|x|^2}{R^2} \right) \right|$$

$$\leq \left\| GQG^T \right\|_\infty \left(\frac{8n|x|^2}{R^3} + \frac{4n}{R} \right) \leq 12n \left\| GQG^T \right\|_\infty,$$

$$(5.7)$$

$$\left| D_w \phi_2 \cdot \nabla \phi_2 \right| = \left| \left(1 - \frac{|x|^2}{R^2} \right)^2 \sum_{i,j=1}^n \left(GQG^T \right)_{ij} \frac{4x_i}{R} \frac{4x_j}{R} \right|$$

$$\leq 16n \left\| GQG^T \right\|_\infty,$$

$$(5.8)$$

and

$$\left| (f - D_w K) \cdot \nabla \phi_2 \right| = \left| (f - D_w K) \frac{4x}{R} \left(1 - \frac{|x|^2}{R^2} \right) \right|$$

$$\leq 4 \left| f - D_w K \right|.$$

$$(5.9)$$

It follows that

$$\sup_{B_R} |\mathrm{I}_1| \leq 14n \left\| GQG^T \right\|_\infty + 4 \left| f - D_w K \right| + N \leq C,$$

by condition (3.5). Similarly,

$$\sup_{B_R} |\mathrm{I}_2| \leq \sup_{B_R} \left[e^{-\sqrt{1+|x|^2}} \left(14n \left\| GQG^T \right\|_\infty \right. \right.$$

$$\left. \left. + 4 \left| f - D_w K \right| \right) \right]$$

$$\leq \tilde{C},$$

by condition (3.6). In the view of Proposition 5.2, one gets

$$\frac{d}{dt} \int_{B_R} \varrho v \leq C \int_{B_R} \varrho v + e^{-R} \tilde{C} \int_{B_R} e^{\sqrt{1+|x|^2}} \rho$$

$$\leq C \int_{B_R} \varrho v + e^{-R+\hat{C}T} \tilde{C} \int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}} \sigma_0(x).$$

$$(5.10)$$

Multiply e^{-Ct} on both sides of (5.10) yields

$$\frac{d}{dt} \left[e^{-Ct} \int_{B_R} \varrho v \right] \leq e^{-R+\hat{C}T-Ct} \tilde{C} \int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}} \sigma_0(x).$$

Integrate from 0 to T and multiply e^{CT} on both sides gives us

$$\int_{B_R} \varrho v(x, T) \leq \|v(x, 0)\|_\infty e^{CT} \int_{B_R} \varrho dx$$

$$+ \frac{e^{CT} - 1}{C} e^{-R+\hat{C}T} \tilde{C} \int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}} \sigma_0(x),$$

where $v(x, 0) = \sigma_0 - \sigma_{0,R}$. Recall that $\varrho(x) = e^{-R\left[-\left(|x|^2/R^2-1\right)^2+1\right]} - e^{-R}$, $|x| \leq R$, we arrive the following estimates:

$$\int_{B_R} \varrho \leq \int_{B_R} \left(1 - e^{-R} \right) \leq C R^n$$

and

$$\int_{B_R} \varrho v(x, T) \geq \int_{B_\frac{R}{2}} \left(e^{-R\left[-\left(|x|^2/R^2-1\right)^2+1\right]} - e^{-R} \right) v(x, T)$$

$$\geq \frac{1}{2} e^{-\frac{7}{16}R} \int_{B_\frac{R}{2}} v(x, T).$$

It is easy to see that $\|v(x, 0)\|_\infty \int_{B_R} \varrho \leq C(n)\epsilon R^n$ is arbitrarily small, since ϵ is independent of R. It yields that

$$\int_{B_\frac{R}{2}} v(x, T) \leq C e^{-\frac{9}{16}R} \int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}} \sigma_0(x), \qquad (1.28)$$

where C is a generic constant, depending on T. ∎

By refining the proofs of Proposition 5.2 and Theorem 3.1, we obtain an interesting property of the density function $\rho(x, t)$. It asserts that ρ captures almost all the density in a large ball. And we could give an precise estimate of the density outside the large ball.

Theorem 5.4: Let $\rho(x, t)$ be a solution of the "pathwise-robust" DMZ equation (2.6) in $\mathbb{Q}_T$. Assume that

1) Condition (3.5) is satisfied;

2) A stronger version of condition (3.6) is valid. To be more precise,

$$e^{-\frac{1}{2}\sqrt{1+|x|^2}} \left[16n \left\| GQG^T \right\|_\infty + 4 \left| f - D_w K \right| \right] \leq C, \tag{5.11}$$

for all $(x,t) \in \mathbb{Q}_T$.

Then

$$\int_{|x| \geq R} \rho(x, T) \leq C e^{-\frac{1}{2}\sqrt{1+R^2}} \int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}} \sigma_0(x), \tag{5.12}$$

where C is a generic constant, which depends on T.

To avoid the distraction, we leave the detailed proof in Appendix B.

B. L^1 convergence

In this section, we shall show that, for any $0 < \tau \leq T$, with the partition $\mathcal{P}_k^\tau = \{0 = \tau_0 < \tau_1 < \cdots < \tau_k = \tau\}$, the L^1 convergence of $\rho_{k,R}(x, \tau)$ to $\rho_R(x, \tau)$ holds, as $k \to +\infty$, where $\rho_{k,R}$ is the solution of (3.11) obtained by our algorithm, and ρ_R is the solution to (3.1). For the clarity, we state the technique lemma will be used in the proof of Theorem 3.2 below.

Lemma 5.2: (Lemma 4.1, [22]) Let Ω be a bounded domain in $\mathbb{R}^n$ and let $v : \overline{\Omega} \times [0, T] \to \mathbb{R}$ be a C^1 function. Assume that $v(x, t) = 0$ for $(x, t) \in \partial\Omega \times [0, T]$. Let $\Omega_t^+ = \{x \in \Omega : v(x,t) \geq 0\}$. Then

$$\frac{d}{dt} \int_{\Omega_t^+} v(x, t) = \int_{\Omega_t^+} \frac{\partial v}{\partial t}(x, t),$$

for almost all $t \in [0, T]$.

Proof of Theorem 3.2: For the notational convenience, we omit the subscript Ω for ρ_Ω and $\rho_{i,\Omega}$ in this proof. Let $\Omega_t^+ = \{x \in \Omega : \rho(x,t) - \rho_i(x,t) \geq 0\}$. Apply Lemma 4.1 to $(\rho - \rho_i)$ taking place of ρ_Ω, with the test function $\psi \equiv 1$, we have

$$\frac{d}{dt} \int_{\Omega_t^+} (\rho - \rho_i) \leq \int_{\Omega_t^+} (\rho - \rho_i) N(\cdot, t)$$
$$+ \int_{\Omega_t^+} \rho_i [N(\cdot, t) - N(\cdot, t; \tau_{i-1})], \tag{5.1}$$

by Lemma 5.2. All the boundary integrals vanish, except $\int_{\partial\Omega_t^+} D_w(\rho - \rho_i) \cdot \nu$, since $(\rho - \rho_i)|_{\partial\Omega_t^+} = 0$. Moreover, $\int_{\partial\Omega_t^+} D_w(\rho - \rho_i) \cdot \nu \leq 0$, due to the similar argument for $\int_{\partial B_R} D_w \rho \cdot \nu \leq 0$ in Proposition 5.2. Combine the conditions (3.8) and (3.9), (5.1) can be controlled by

$$\frac{d}{dt} \int_{\Omega_t^+} (\rho - \rho_i) \leq C \int_{\Omega_t^+} (\rho - \rho_i) + \tilde{C}(t - \tau_{i-1})^\alpha \int_\Omega \rho. \tag{5.2}$$

To estimate $\int_\Omega \rho$, we apply Lemma 4.1 to ρ, with the test function $\psi \equiv 1$, we get

$$\frac{d}{dt} \int_\Omega \rho \leq \int_\Omega \rho N \leq C \int_\Omega \rho,$$

which implies

$$\int_\Omega \rho \leq C \int_\Omega \sigma_{0,\Omega}, \tag{5.3}$$

where C is a generic constant, depending on T, for all $0 \leq t \leq T$. Thus,

$$\frac{d}{dt} \int_{\Omega_t^+} (\rho - \rho_i) \leq C \int_{\Omega_t^+} (\rho - \rho_i) + \tilde{C}(t - \tau_{i-1})^\alpha \int_\Omega \sigma_{0,\Omega}.$$

Multiply $e^{-\tilde{C}(t-\tau_{i-1})}$ on both sides and integrate from τ_{i-1} to t, we get

$$\int_{\Omega_t^+} (\rho - \rho_i)(x, t) \leq e^{\tilde{C}(t-\tau_{i-1})} \int_{\Omega_{\tau_{i-1}}^+} (\rho - \rho_i)(x, \tau_{i-1})$$
$$+ C \frac{(t - \tau_{i-1})^{1+\alpha}}{1 + \alpha} e^{\tilde{C}(t-\tau_{i-1})},$$

where C is a constant, which depends on T, $\int_\Omega \sigma_{0,\Omega}$. Similarly, one can also get, for $\Omega_t^- = \{x \in \Omega : \rho(x,t) - \rho_i(x,t) < 0\}$, that

$$\int_{\Omega_t^-} (\rho_i - \rho)(x, t) \leq e^{\tilde{C}(t-\tau_{i-1})} \int_{\Omega_{\tau_{i-1}}^-} (\rho_i - \rho)(x, \tau_{i-1})$$
$$+ C \frac{(t - \tau_{i-1})^{1+\alpha}}{1 + \alpha} e^{\tilde{C}(t-\tau_{i-1})}.$$

Consequently, we have

$$\int_\Omega |\rho - \rho_i|(x, t)$$
$$\leq e^{\tilde{C}(t-\tau_{i-1})} \left[\int_\Omega |\rho - \rho_i|(x, \tau_{i-1}) + C \frac{(t - \tau_{i-1})^{1+\alpha}}{1 + \alpha} \right]$$
$$\leq e^{\tilde{C}(t-\tau_{i-1})} \left[\int_\Omega |\rho - \rho_{i-1}|(x, \tau_{i-1}) + C \frac{(t - \tau_{i-1})^{1+\alpha}}{1 + \alpha} \right], \tag{5.4}$$

since $\rho_i(x, \tau_{i-1}) = \rho_{i-1}(x, \tau_{i-1})$, for $i = 1, 2, \cdots, k$. Applying (5.4) recursively, we obtain

$$\int_\Omega |\rho - \rho_k|(x, \tau_k)$$
$$\leq e^{\tilde{C}(\tau_k - \tau_{k-1})} \left[\int_\Omega |\rho - \rho_{k-1}|(x, \tau_k) + C \frac{(\tau_k - \tau_{k-1})^{1+\alpha}}{1 + \alpha} \right]$$
$$\leq e^{\tilde{C}T} \int_\Omega |\rho - \rho_0|(x, 0)$$
$$+ \frac{C}{1+\alpha} [(\tau_k - \tau_{k-1})^{1+\alpha} e^{\tilde{C}(\tau_k - \tau_{k-1})}$$
$$+ (\tau_{k-1} - \tau_{k-2})^{1+\alpha} e^{\tilde{C}(\tau_k - \tau_{k-2})}$$
$$+ \cdots + (\tau_1 - \tau_0)^{1+\alpha} e^{\tilde{C}(\tau_k - \tau_0)}]$$
$$= \frac{C}{1+\alpha} \frac{T^{1+\alpha}}{k^{1+\alpha}} \left(e^{\tilde{C}\frac{T}{k}} + e^{\tilde{C}\frac{2T}{k}} + \cdots + e^{\tilde{C}\frac{kT}{k}} \right) \leq \frac{C}{k^\alpha},$$

where C is a constant, which depends on α, T and $\int_\Omega \sigma_{0,\Omega}$. It is clear that $\int_\Omega |\rho - \rho_k| \to 0$, as $k \to \infty$. ∎

VI. LOWER BOUND ESTIMATE OF DENSITY FUNCTION

It is well-known that solving the "pathwise-robust" DMZ equation numerically is not easy because it is easily vanishing. We are also interested in whether the lower bound of the density function could be derived in the case where the drift term f and the observation term h are with at most

the polynomial growth. The theorem below gives this lower bound:

Theorem 6.5: Let ρ_R be the solution of (3.1), the "pathwise-robust" DMZ equation on B_R. Assume that

1) $f(x,t)$ and $h(x,t)$ have at most polynomial growth in $|x|$, for all $t \in [0, T]$;
2) For any $0 \le t \le T$, there exists positive integer m and positive constants C' and C'' independent of R such that the following two conditions hold on $\mathbb{R}^n$:

(a)
$$\frac{|x|^{m-2}}{2} \left[nm(m-2) \left\| GQG^T \right\|_\infty \right.$$
$$\left. + m \operatorname{Tr} \left(GQG^T \right) \right]$$
$$- m|x|^{m-2}(f - D_w K) \cdot x + N(x,t) \ge -C'; \tag{6.1}$$

(b)
$$\left| n \left\| GQG^T \right\|_\infty \left(\frac{1}{2} m^2 |x|^{2m-2} \right. \right.$$
$$\left. - m \left(\frac{1}{2} m - 1 \right) |x|^{m-2} \right)$$
$$- \frac{1}{2} m \operatorname{Tr} \left(GQG^T \right) |x|^{m-2}$$
$$\left. - m(f - D_w K) \cdot x |x|^{m-2} \right|$$
$$\le \frac{1}{2} nm(m+1) \left\| GQG^T \right\|_\infty |x|^{2m-2} + C'', \tag{6.2}$$

where $\operatorname{Tr}(*)$ is the trace of $*$.
3) Condition (3.8) is satisfied.

Then for any $R_0 < R$,

$$\int_{B_{R_0}} \zeta \rho_R(x, T)$$
$$\ge \frac{e^{(C-C')T - R_0^m}}{C'} \left(\frac{1}{2} nm(m+1) \left\| GQG^T \right\|_\infty R_0^{2m-2} + C'' \right)$$
$$\cdot \left(1 - e^{C'T} \right) \int_{B_R} \sigma_{0,R}(x) + e^{-C'T} \int_{B_{R_0}} \zeta \sigma_{0,R}(x),$$

where $\zeta(x) = e^{-\xi(x)} - e^{-\xi(R_0)}$, $\xi(x) = |x|^m$.

In particular, the solution ρ of the "pathwise-robust" DMZ equation (2.5) on $\mathbb{R}^n$ has the estimate

$$\int_{\mathbb{R}^n} e^{-|x|^m} \rho(x, T) \ge e^{-C'T} \int_{\mathbb{R}^n} e^{-|x|^m} \sigma_0(x).$$

Proof: Apply Lemma 4.1 to ρ_R with the test function ψ to be $\zeta = e^{-\xi(x)} - e^{-\xi(R_0)}$, where $\xi(x)$ is an increasing function in $|x|$, we have

$$\frac{d}{dt} \int_{B_{R_0}} \zeta \rho_R = \int_{B_{R_0}} \rho_R \left[\frac{1}{2} D_w^2 \zeta + (f - D_w K) \cdot \nabla \zeta + \zeta N \right].$$

All the boundary integrals vanish, since $\zeta|_{\partial B_R} = \rho_R|_{\partial B_R} = 0$. Direct computations yield that

$$\frac{d}{dt} \int_{B_{R_0}} \zeta \rho_R$$
$$= \int_{B_{R_0}} \rho_R e^{-\xi(R_0)}$$
$$\cdot \left\{ \frac{1}{2} \frac{\xi'^2(r)}{r^2} \sum_{i,j=1}^n \left(GQG^T \right)_{ij} x_i x_j - \frac{\xi'(r)}{r} (f - D_w K) \cdot x \right.$$
$$- \frac{1}{2} \sum_{i,j=1}^n \left(GQG^T \right)_{ij} \left[\left(\xi''(r) - \frac{\xi'(r)}{r} \right) \frac{x_i x_j}{r^2} \right]$$
$$\left. - \frac{1}{2} \operatorname{Tr} \left(GQG^T \right) \frac{\xi'(r)}{r} \right\}$$
$$+ \int_{B_{R_0}} \zeta \rho_R \left[\frac{1}{2} D_w \xi \cdot \nabla \xi - \frac{1}{2} D_w^2 \xi - (f - D_w K) \cdot \nabla \xi + N \right]$$
$$\triangleq I_3 + \int_{B_{R_0}} \zeta \rho_R [I_4].$$

Let $\xi(r) = r^m$, where $r = |x|$, m is some positive integer sufficiently large. Through elementary computations, we get

$$I_4$$
$$= \frac{1}{2} \frac{\xi'^2(r)}{r^2} \sum_{i,j=1}^n \left(GQG^T \right)_{ij} x_i x_j$$
$$- \frac{1}{2} \left[m(m-2) r^{m-4} \sum_{i,j=1}^n \left(GQG^T \right)_{ij} x_i x_j \right.$$
$$\left. + m r^{m-2} \operatorname{Tr} \left(GQG^T \right) \right] - m r^{m-2}(f - D_w K) \cdot x + N$$
$$\ge -\frac{1}{2} \left[nm(m-2) \left\| GQG^T \right\|_\infty + m \operatorname{Tr} \left(GQG^T \right) \right] r^{m-2}$$
$$- m r^{m-2}(f - D_w K) \cdot x + N \ge C',$$

where C' is a positive constant independent of R_0, by condition (6.1). For large enough m, we have

$$|I_3| \le e^{-R_0^m}$$
$$\cdot \int_{B_R} \left| n \left\| GQG^T \right\|_\infty \left[\frac{1}{2} m^2 r^{2m-2} - m \left(\frac{1}{2} m - 1 \right) r^{m-2} \right] \right.$$
$$\left. - \frac{1}{2} m \operatorname{Tr} \left(GQG^T \right) r^{m-2} - m(f - D_w K) \cdot x r^{m-2} \right| \rho_R$$
$$\le e^{-R_0^m} \left(\frac{1}{2} nm(m+1) \left\| GQG^T \right\|_\infty R_0^{2m-2} + C'' \right) \int_{B_R} \rho_R$$
$$\le \left(\frac{1}{2} nm(m+1) \left\| GQG^T \right\|_\infty R_0^{2m-2} + C'' \right)$$
$$\cdot e^{CT - R_0^m} \int_{B_R} \sigma_{0,R} \triangleq \gamma(R_0).$$

The last inequality follows by the similar argument of (5.3). Hence,

$$\frac{d}{dt} \int_{B_{R_0}} \zeta \rho_R \ge -\gamma(R_0) - C' \int_{B_{R_0}} \zeta \rho_R.$$

This implies

$$\int_{B_{R_0}} \zeta \rho_R(x, T)$$

$$\geq e^{-C'T} \int_{B_{R_0}} \zeta \sigma_{0,R}(x) + \frac{\gamma(R_0)}{C'} \left(e^{-C'T} - 1 \right)$$

$$\geq e^{-C'T} \int_{B_{R_0}} \zeta \sigma_{0,R}(x)$$

$$+ \left(\frac{1}{2} nm(m+1) \left| \left| GQG^T \right| \right|_\infty R_0^{2m-2} + C'' \right)$$

$$\cdot \frac{e^{(C-C')T - R_0^m}}{C'} \left(1 - e^{C'T} \right) \int_{B_R} \sigma_{0,R}(x). \quad (6.3)$$

Let $R_0 \to \infty$, we have

$$\int_{\mathbb{R}^n} e^{-|x|^m} \rho(x, T) \geq e^{-C'T} \int_{\mathbb{R}^n} e^{-|x|^m} \sigma_0(x).$$

∎

VII. NUMERICAL SIMULATIONS

In this section, we shall apply our algorithm to both "time-invariant" case and "time-varying" case. The numerical simulations support our theorems. In our implementation, we adopt the Hermite spectral method (HSM) to get the approximate solution of (2.13). Thus, the basis functions $\{\phi_l\}_{l=1}^\infty$ in (2.14) are choosen to be the generalized Hermite functions $\{H_n^{\alpha,\beta}(x)\}_{n=0}^\infty$. We refer the interested readers to the detailed definitions in [14].

For $N > 0$, let us denote $\mathcal{R}_N$ the subspace spanned by the first N generalized Hermite functions:

$$\mathcal{R}_N = \text{span}\{H_0^{\alpha,\beta}(x), \cdots, H_N^{\alpha,\beta}(x)\}.$$

The formulation of HSM to (2.13) in 1-dimension is to find $u_N(x, t) \in \mathcal{R}_N$ such that

$$\begin{cases} \langle \partial_t u_N(x, t), \varphi \rangle = -\frac{1}{2} \langle \partial_x[(GQG^T)u_N], \partial_x \varphi \rangle \\ \qquad\qquad + \langle f u_N, \partial_x \varphi \rangle - \frac{1}{2} \langle (h^T S^{-1} h) u_N, \varphi \rangle \\ u_N(x, 0) = P_N u_0(x), \end{cases}$$
$$(7.4)$$

for any $\varphi \in \mathcal{R}_N$, where $\langle \cdot, \cdot \rangle$ denotes the scalar product in $L^2(\mathbb{R})$ and P_N is the projection operator such that $P_N : L^2(\mathbb{R}) \to \mathcal{R}_N$. Write the solution $u_N \in \mathcal{R}_N$ in the form

$$u_N(x, t) = \sum_{n=0}^N a_n(t) H_n^{\alpha,\beta}(x),$$

and take the test function $\varphi \in \mathcal{R}_N$ in (7.4) to be $H_n^{\alpha,\beta}$, $n = 0, \cdots, N$. From (7.4) and the properties of generalized Hermite functions, $\vec{a}(t) := (a_0(t), a_1(t), \cdots, a_N(t))^T$ satisfies the ODE

$$\partial_t \vec{a}(t) = A \vec{a}(t), \quad (7.5)$$

where A is a $(N+1) \times (N+1)$ matrix, may depend on t, if G, Q, f or h is explicitly time-dependent. This ODE can be precomputed. The only difference between "time-varying" case and "time-invariant" case is that it costs much more

Fig. 1. 1D cubic sensor, with the initial condition $u_0(x) = e^{-\frac{x^4}{4}}$. Black: real state; Green: extended Kalman filter; Red: our algorithm.

memory to store the off-line data in the "time-varying" case, as we explained before in section II. Nevertheless, it doesn't change the off-line virture of our algorithm. We refer the interested readers of the implementation to [14], and we shall omit the technical details in this paper.

Once u_N at each step is obtained, $\hat{\rho}$ can be recovered by (2.12). The conditional expectation of the state x_t is computed by definition

$$\mathbb{E}[x, \{y_\tau\}_{0 \leq \tau \leq t}](t) = \frac{\int_\mathbb{R} x \hat{\rho}(x, t) dx}{\int_\mathbb{R} \hat{\rho}(x, t) dx}.$$

A. "time-invariant" case: the 1D cubic sensor

Let us consider the following model

$$\begin{cases} dx_t = dv_t \\ dy_t = x_t^3 dt + dw_t, \end{cases}$$

where $x_t, y_t \in \mathbb{R}$, v_t, w_t are scalar Brownian motion processes with $E[dv_t^T dv_t] = 1$, $E[dw_t^T dw_t] = 1$. The 1D Kolmogorov forward equation (2.13) here is

$$u_t = \frac{1}{2} u_{xx} - \frac{1}{2} x^6 u, \quad (7.6)$$

at each time step. We assume the inital density function $u_0(x) = e^{-x^4/4}$ and the updated initial data are

$$u_i(x, \tau_i) = e^{x^3 \cdot dy_t} u_{i-1}(x, \tau_i).$$

In Figure 1, we see that our algorithm tracks the state's expectation very well, while the extended Kalman filter (EKF) completely fails around $t = 5$. The total simulation time is $T = 50$, and the update time step is $dt = \tau_{i+1} - \tau_i = 0.01$. It costs our algorithm only around $4.88s$ to finish the simulation, i.e. the updated time is less than $10^{-3}s$.

Fig. 2. 1D "time-varying" almost linear sensor, with the initial condition $u_0(x) = e^{-\frac{x^2}{2}}$. Black: real state; Green: extended Kalman filter; Red: our algorithm.

B. "time-varying" case: the 1D almost linear sensor

The 1D almost linear sensor we are considering is

$$\begin{cases} dx_t = [1 + 0.1\cos(20\pi t)]dv_t \\ dy_t = x_t[1 + 0.25\cos(x_t)]dt + dw_t, \end{cases} \quad (7.7)$$

where $x_t, y_t \in \mathbb{R}$, v_t, w_t are scalar Brownian motion processes with $E[dv_t^T dv_t] = E[dw_t^T dw_t] = 1$. The Kolmogorov forward equation (2.13) in this example is

$$u_t = \frac{1}{2}[1 + 0.1\cos(20\pi t)]^2 u_{xx} - \frac{1}{2}x^2[1 + 0.25\cos(x)]^2 u,$$

with the initial data $u_0(x) = e^{-x^2/2}$ and the updated initial data

$$u_i(x, \tau_i) = e^{x^2[1 + 0.25\cos(x)] \cdot dy_t} u_{i-1}(x, \tau_i),$$

$i = 1, 2, \cdots, k$. In Figure 2, our algorithm tracks the state's expectation at least as well as the EKF. The total simulation time is $T = 60$, and the update time step is $dt = \tau_{i+1} - \tau_i = 0.01$. It costs our algorithm only around $3.17s$ to complete the simulatoin, i.e. the updated time is less than $5 \times 10^{-4}s$.

VIII. CONCLUSION

In this paper, we extend the algorithm developed in [22] to the most general nonlinear filterings. We theoretically verified that under very mild growth conditions on the drift term and the observation term, the unique non-negative weak solution ρ of its associated "pathwise-robust" DMZ equation can be approximated by the solution ρ_R of the DMZ equation restricted on a large ball B_R with 0-Dirichlet boundary condition. The error of this approximation tends to zero exponentially as the radius of the ball R approaching infinity. Moreover, ρ_R can be efficiently approximated by our algorithm. We show that the approximate solution $\hat{\rho}_R$ obtained by our algorithm converges to ρ_R in the L^1 sense for all $t \in [0, T]$, as the partition of time becomes finer, and a precise error estimate of this convergence is given explicitly. Equally important, our algorithm preserves the two advantages of KF: "memoryless" and "real time". We

also give the detail explanation of the off-line virture of our algorithm in the formulation. Numerical experiments support the feasibility and efficiency of our algorithm.

APPENDIX A
EXISTENCE AND UNIQUENESS OF THE SOLUTION

Before we show the existence of the weak solution, we shall give a priori estimations of up to the first order derivative of the solution to the robust DMZ equation on $B_R \times [0, T]$.

Theorem A.1: Consider the "pathwise-robust" DMZ equation (3.1) on $\mathbb{Q}_R := B_R \times [0, T]$, where $B_R = \{x \in \mathbb{R}^n : |x| \leq R\}$ is a ball of radius R. Assume that

$$\left\| \frac{d}{dt}(GQG^T) \right\|_\infty < \infty, \quad (A.1)$$

for all $t \in [0, T]$. Suppose there exists a positive function $g(x)$ on $\mathbb{R}^n$ such that for all $t \in [0, T]$, g and $\tilde{g} \triangleq g + \log|D_w J|$ satisfy

1) $\left| D_w g + \frac{1}{2}\nabla(GQG^T) - F \right|^2 + 2\lambda_1 J \leq C$, (A.2)
2) $D_w^2 g + 2D_w g \cdot \nabla g + 2[\nabla(GQG^T) - F] \cdot \nabla g + \frac{1}{2}\nabla^2(GQG^T) - \text{div}F + J \leq C$, (A.3)
3) $D_w^2 \tilde{g} + 2D_w \tilde{g} \cdot \nabla \tilde{g} + 2[\nabla(GQG^T) - F] \cdot \nabla \tilde{g} + \frac{1}{2}\nabla^2(GQG^T) - \text{div}F + J \leq C$, (A.4)
4) $\int_{\mathbb{R}^n} e^{2\tilde{g}}\sigma^2(x) \leq C$ and $\int_{\mathbb{R}^n} e^{2g}D_w\sigma \cdot \nabla\sigma \leq C$,(A.5)

where C is a generic constant, which may differ from line to line, and $\nabla(*) = \left[\sum_{i=1}^n \frac{\partial(*)_{ij}}{\partial x_i}\right]_{j=1}^n$, $\nabla^2(*) = \sum_{i,j=1}^n \frac{\partial^2(*)_{ij}}{\partial x_i \partial x_j}$. Then, for $0 \leq t \leq T$,

$$\int_{B_R} e^{2g}\rho_R^2(x,t)dx \leq e^{Ct}\int_{B_R} e^{2g}\sigma^2(x)dx, \quad (A.6)$$

$$\int_{B_R} e^{2g}D_w\rho_R(x,t) \cdot \nabla\rho_R(x,t)dx$$
$$\leq e^{Ct}\int_{B_R} e^{2g}D_w\sigma(x) \cdot \nabla\sigma(x)dx + Ce^{Ct}\int_{B_R} e^{2\tilde{g}}\sigma^2(x)dx, \quad (A.7)$$

where D_w and $J(x,t)$ is defined in (3.4)and (2.9), respectively.

Remark A.1: The conditions in Theorem A.1 are easily checked, if the drift terms $h(x)$ and $f(x)$ are at most polynomial growth in $r = |x|$. However, in general, the existence of such g is not always available.

Proof: Let g be some positive function on $\mathbb{R}^n$.

$$\frac{d}{dt}\int_{B_R} e^{2g}\rho_R^2 = \int_{B_R} e^{2g}\rho_R D_w^2 \rho_R + 2\int_{B_R} e^{2g}\rho_R(F \cdot \rho_R)$$
$$+ 2\int_{B_R} e^{2g}J\rho_R^2$$
$$\triangleq \text{I} + \text{II} + \text{III}. \quad (A.8)$$

Apply integration by parts to I and II in (A.8)

$$\text{I} = -2\int_{B_R} \rho_R e^{2g}D_w g \cdot \nabla\rho_R - \int_{B_R} e^{2g}D_w\rho_R \cdot \nabla\rho_R$$
$$- \int_{B_R} e^{2g}\rho_R\nabla(GQG^T) \cdot \nabla\rho_R$$
$$\leq -2\int_{B_R} \rho_R e^{2g}D_w g \cdot \nabla\rho_R - \int_{B_R} e^{2g}\rho_R\nabla(GQG^T) \cdot \nabla\rho_R$$
$$\triangleq \text{I}_1 + \text{I}_2.$$

Integration by parts further, we have

$$
\begin{aligned}
I_1 =& 4\int_{B_R} e^{2g}\rho_R^2 D_w g \cdot \nabla g + 2\int_{B_R} e^{2g}\rho_R D_w g \cdot \nabla \rho_R \\
&+ 2\int_{B_R} e^{2g}\rho_R^2 \nabla(GQG^T)\cdot \nabla g + 2\int_{B_R} e^{2g}\rho_R^2 D_w^2 g.
\end{aligned}
\tag{A.9}
$$

Notice that the second term of the right-hand side of (A.9) is $-I_1$, we have

$$
\begin{aligned}
I_1 =& 2\int_{B_R} e^{2g}\rho_R^2 D_w g \cdot \nabla g \\
&+ \int_{B_R} e^{2g}\rho_R^2[\nabla(GQG^T)\cdot\nabla g + D_w^2 g].
\end{aligned}
\tag{A.10}
$$

The similar argument applies to I_2:

$$
\begin{aligned}
I_2 =& \int_{B_R} e^{2g}\rho_R^2 \nabla(GQG^T)\cdot \nabla g \\
&+ \frac{1}{2}\int_{B_R} e^{2g}\rho_R^2 \nabla^2(GQG^T).
\end{aligned}
\tag{A.11}
$$

Thus,

$$
\begin{aligned}
I \le& \int_{B_R} e^{2g}\rho_R^2 \Big[D_w^2 g + 2D_w g \cdot \nabla g \\
&+ 2\nabla(GQG^T)\cdot \nabla g + \frac{1}{2}\nabla^2(GQG^T)\Big].
\end{aligned}
\tag{A.12}
\tag{A.13}
$$

The same trick of I_1 applies to II in (A.8), we obtain

$$
II = -\int_{B_R} e^{2g}\rho_R^2[2F\cdot\nabla g + \mathrm{div}F].
\tag{A.14}
$$

Substitute (A.12) and (A.14) back to (A.8), we obtain

$$
\begin{aligned}
&\frac{d}{dt}\int_{B_R} e^{2g}\rho_R^2 \\
&\le \int_{B_R} e^{2g}\rho_R^2\Big\{ D_w^2 g + 2D_w g\cdot\nabla g + 2[\nabla(GQG^T)-F]\cdot\nabla g \\
&\quad + \frac{1}{2}\nabla^2(GQG^T) - \mathrm{div}F + J\Big\} \\
&\le C\int_{B_R} e^{2g}\rho_R^2,
\end{aligned}
$$

by condition (A.3). (A.6) follows directly from Gronwall's inequality. To show (A.7), we consider

$$
\begin{aligned}
&\frac{d}{dt}\int_{B_R} e^{2g} D_w \rho_R \cdot \nabla \rho_R \\
&= \int_{B_R} e^{2g}\sum_{i,j=1}^n \frac{d}{dt}(GQG^T)_{ij}\frac{\partial\rho_R}{\partial x_i}\frac{\partial\rho_R}{\partial x_j} \\
&\quad + 2\int_{B_R} e^{2g}\sum_{i,j=1}^n (GQG^T)_{ij}\frac{\partial}{\partial x_i}\left(\frac{\partial\rho_R}{\partial t}\right)\frac{\partial\rho_R}{\partial x_j} \\
&\triangleq IV + V.
\end{aligned}
\tag{A.15}
$$

Due to condition (A.1), IV of (A.15) turns out to be

$$
\begin{aligned}
IV \le& \frac{1}{2}\left\|\frac{d}{dt}(GQG^T)\right\|_\infty \int_{B_R} e^{2g}\sum_{i,j=1}^n\left[\left(\frac{\partial\rho_R}{\partial x_i}\right)^2 + \left(\frac{\partial\rho_R}{\partial x_j}\right)^2\right] \\
=& n\left\|\frac{d}{dt}(GQG^T)\right\|_\infty \int_{B_R} e^{2g}|\nabla\rho_R|^2 \\
\le& \frac{n}{\lambda_1}\left\|\frac{d}{dt}(GQG^T)\right\|_\infty \int_{B_R} e^{2g} D_w\rho_R\cdot\nabla\rho_R,
\end{aligned}
\tag{A.16}
$$

since $D_w\rho_R\cdot\nabla\rho_R \ge \lambda_1|\nabla\rho_R|^2$. Next, V in (A.15) is

$$
\begin{aligned}
V =& -2\int_{B_R} e^{2g}[(2D_w g + \nabla(GQG^T))\cdot\nabla\rho_R + D_w^2\rho_R] \\
&\cdot\left(\frac{1}{2}D_w^2\rho_R + F\cdot\nabla\rho_R + J\rho_R\right) \\
=& -\int_{B_R} e^{2g}\left\{ D_w^2\rho_R + \left[D_w g + \frac{1}{2}\nabla(GQG^T)+F\right]\cdot\nabla\rho_R\right\}^2 \\
&+ \int_{B_R} e^{2g}\left[D_w g + \frac{1}{2}\nabla(GQG^T) - F\right]^2|\nabla\rho_R|^2 \\
&- 2\int_{B_R} e^{2g}[D_w^2\rho_R + (2D_w g + \nabla(GQG^T))\cdot\nabla\rho_R]J\rho_R \\
\le& \int_{B_R} e^{2g}\left[D_w g + \frac{1}{2}\nabla(GQG^T) - F\right]^2|\nabla\rho_R|^2 \\
&- 2\int_{B_R} e^{2g}[D_w^2\rho_R + (2D_w g + \nabla(GQG^T))\cdot\nabla\rho_R]J\rho_R.
\end{aligned}
\tag{A.17}
$$

Notice that

$$
\begin{aligned}
&\int_{B_R} e^{2g} D_w^2\rho_R J\rho_R \\
&= -\int_{B_R} e^{2g}\big[2(D_w g\cdot\nabla\rho_R)J\rho_R + JD_w\rho_R\cdot\nabla\rho_R \\
&\quad + (D_w\rho_R\cdot\nabla J)\rho_R + \nabla(GQG^T)\cdot\nabla\rho_R J\rho_R\big].
\end{aligned}
\tag{A.18}
$$

Take (A.18) into account, V becomes

$$
\begin{aligned}
V \le& \int_{B_R} e^{2g} \\
&\cdot\left\{ \frac{1}{\lambda_1}\left\{\left[D_w g + \frac{1}{2}\nabla(GQG^T) - F\right]^2 + 1\right\} + 2J\right\} \\
&\cdot D_w\rho_R\cdot\nabla\rho_R + \int_{B_R} e^{2g}|D_w J|^2\rho_R^2.
\end{aligned}
\tag{A.19}
$$

Combine (A.16) and (A.19), we have

$$
\begin{aligned}
&\frac{d}{dt}\int_{B_R} e^{2g} D_w\rho_R\cdot\nabla\rho_R \\
&\le \int_{B_R} e^{2g}\left\{ \frac{1}{\lambda_1}\left\{ n\left\|\frac{d}{dt}(GQG^T)\right\|_\infty\right.\right. \\
&\quad \left.\left. + \left[D_w g + \frac{1}{2}\nabla(GQG^T) - F\right]^2 + 1\right\} + 2J\right\} \\
&\quad \cdot D_w\rho_R\cdot\nabla\rho_R \\
&\quad + \int_{B_R} e^{2g}|D_w J|^2\rho_R^2.
\end{aligned}
\tag{A.20}
$$

By conditions (A.2)-(A.5), the estimate (A.7) follows immediately. $\qquad\square$

Proof of existence in Theorem 4.3: Let R_k be a sequence of positive number such that $\lim_{k\to\infty} R_k = \infty$. Let $\rho_k(x,t)$ be the solution of the "pathwise-robust" DMZ equation (3.1) on $B_{R_k} \times [0,T]$, where $B_{R_k} = \{x \in \mathbb{R}^n : |x| \le R_k\}$ is a ball of radius R_k. In view of Theorem A.1, the sequence $\{\rho_k\}$ is a bounded set in $H_0^{1;1}(\mathbb{Q}_{R_k})$. Thus, there exists a subsequence $\{\rho_{k'}\}$ which is weakly convergent to ρ. Moreover, ρ has the weak derivative $\frac{\partial \rho}{\partial x_i} \in L^2(\mathbb{Q}_{R_k})$, and $\frac{\partial \rho_{k'}}{\partial x_i}$ weakly tends to it. Now we claim that the weak derivative $\frac{\partial \rho}{\partial t}$ exists. To see this, let $\Phi(x,t) \in H_0^{1;1}(\mathbb{Q}_{R_k})$, then

$$
\iint_{\mathbb{Q}_{R_k}} \frac{1}{2} \sum_{i,j=1}^{n} (GQG^T)_{ij} \frac{\partial \Phi}{\partial x_j} \frac{\partial \rho}{\partial x_i}
$$
$$
+ \left[\sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{\partial (GQG^T)_{ij}}{\partial x_j} - F_i \right) \frac{\partial \rho}{\partial x_i} - J\rho \right] \Phi
$$
$$
= \lim_{k'\to\infty} \iint_{\mathbb{Q}_{R_k}} \frac{1}{2} \sum_{i,j=1}^{n} (GQG^T)_{ij} \frac{\partial \Phi}{\partial x_j} \frac{\partial \rho_{k'}}{\partial x_i}
$$
$$
+ \left[\sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{\partial (GQG^T)_{ij}}{\partial x_j} - F_i \right) \frac{\partial \rho_{k'}}{\partial x_i} - J\rho_{k'} \right] \Phi
$$
$$
= - \lim_{k'\to\infty} \iint_{\mathbb{Q}_{R_k}} \frac{\partial \rho_{k'}}{\partial t} \Phi = \lim_{k'\to\infty} \iint_{\mathbb{Q}_{R_k}} \rho_{k'} \frac{\partial \Phi}{\partial t}
$$
$$
= \iint_{\mathbb{Q}_{R_k}} \rho \frac{\partial \Phi}{\partial t}.
$$

Clearly, $\rho(x,0) = \lim_{k'\to\infty} \rho_{k'}(x,0) = \sigma_0(x)$.

Theorem A.2: Assume further that for some $c > 0$,

$$
\sup_{0\le t\le T} \int_{\mathbb{R}^n} e^{cr} \rho^2(x,t) dx < \infty, \qquad (A.21)
$$

and

$$
\int_{\mathbb{Q}_T} |\nabla \rho(x,t)|^2 dx dt < \infty, \qquad (A.22)
$$

where $r = |x|$. Suppose that there exists a finite number $\alpha > 0$ such that

$$
2J(x,t) - \frac{1}{4\lambda_1}[cD_w r - (F(x,t) + \tilde{F}(x,t))]^2 \le \alpha, \quad (A.23)
$$

for all $(x,t) \in \mathbb{Q}_T$, where λ_1 is the smallest eigenvalue of the matrix (GQG^T),

$$
\tilde{F}(x,t) = \left[\frac{1}{2} \sum_{j=1}^{n} (GQG^T)_{ij} + \sum_{j=1}^{n} (GQG^T)_{ij} \frac{\partial K}{\partial x_j} - f_i \right]_{i=1}^{n}, \quad (A.24)
$$

and $J(x,t)$ is defined as in (2.9). Then the non-negative weak solution $\rho(x,t)$ of the "pathwise-robust" DMZ equation on $\mathbb{Q}_T$ is unique.

Proof of uniqueness of Theorem 4.3 (Theorem A.2): To show the uniqueness of the solution, we only need to show that $\rho(x,t) = 0$ on $\mathbb{Q}_T$ if $\rho(x,0) = 0$. Let $\alpha T < 1$. For any

test function $\psi(x,t) = e^{cr}\Phi(x,t)$, where $r = |x|$, c is some constant and $\Phi(x,t) \in H_0^{1;1}(\mathbb{Q}_T)$, then $\rho(x,t)$ satisfies

$$
\int_{\mathbb{R}^n} \rho(x,T)\Phi(x,T)e^{cr}dx - \int_0^T \int_{\mathbb{R}^n} \rho(x,t)\frac{\partial \Phi}{\partial t}(x,t)e^{cr}dxdt
$$
$$
= \int_{\mathbb{Q}_T} -\frac{1}{2}e^{cr}\nabla\Phi(x,t) \cdot D_w\rho(x,t) - \frac{c}{2}e^{cr}\Phi(x,t)\nabla r \cdot D_w\rho(x,t)
$$
$$
+ \tilde{F}(x,t) \cdot \nabla\rho(x,t)\Phi(x,t)e^{cr}
$$
$$
+ J(x,t)\rho(x,t)\Phi(x,t)e^{cr}dxdt. \qquad (A.25)
$$

where $\tilde{F}$ is defined in (A.24). Approximate $\rho(x,t)$ by $\Phi(x,t)$ in the $H^{1;1}(\mathbb{Q}_T)$-norm, we get

$$
\int_{\mathbb{R}^n} \rho^2(x,T)e^{cr}dx
$$
$$
= \int_{\mathbb{Q}_T} e^{cr}\left[-D_w\rho(x,t) \cdot \nabla\rho(x,t) - c\rho(x,t)\nabla r \cdot D_w\rho(x,t) \right.
$$
$$
+ (\tilde{F}(x,t) + F(x,t)) \cdot \nabla\rho(x,t)\rho(x,t)
$$
$$
\left. + 2J(x,t)\rho^2(x,t) \right] dxdt.
$$
$$
\le \int_{\mathbb{Q}_T} e^{cr}[-\lambda_1|\nabla\rho(x,t)|^2 - c\rho(x,t)D_w r \cdot \nabla\rho(x,t)
$$
$$
+ (F(x,t) + \tilde{F}(x,t)) \cdot \nabla\rho(x,t)\rho(x,t)
$$
$$
+ 2J(x,t)\rho^2(x,t)]dxdt.
$$
$$
= -\lambda_1 \int_{\mathbb{Q}_T} e^{cr} \left\{ \frac{1}{2\lambda_1}[cD_w r - (F(x,t) + \tilde{F}(x,t))]\rho(x,t) \right.
$$
$$
\left. + |\nabla\rho(x,t)| \right\}^2 dxdt
$$
$$
+ \int_{\mathbb{Q}_T} e^{cr} \left\{ 2J(x,t) - \frac{1}{4\lambda_1}[cD_w r - (F(x,t) + \tilde{F}(x,t))]^2 \right\}
$$
$$
\cdot \rho^2(x,t)dxdt
$$
$$
\le \int_{\mathbb{Q}_T} e^{cr} \left\{ 2J(x,t) - \frac{1}{4\lambda_1}[cD_w r - (F(x,t) + \tilde{F}(x,t))]^2 \right\}
$$
$$
\cdot \rho^2(x,t)dxdt, \qquad (A.26)
$$

due to the positive definite of (GQG^T). By condition (A.23), we have

$$
\int_{\mathbb{R}^n} e^{cr}\rho^2(x,T)dx \le \alpha \int_{\mathbb{Q}_T} e^{cr}\rho^2(x,t)dxdt. \qquad (A.27)
$$

According to the mean value theorem, there exists $T_1 \in (0,T)$ such that

$$
\int_{\mathbb{Q}_T} e^{cr}\rho^2(x,t)dxdt = \int_0^T \int_{\mathbb{R}^n} e^{cr}\rho^2(x,t)dxdt
$$
$$
= T \int_{\mathbb{R}^n} e^{cr}\rho^2(x,T_1)dx. \qquad (A.28)
$$

Apply (A.27) and (A.28) recursively, there exists $T_m \in (0, T)$ such that

$$\int_{\mathbb{R}^n} e^{cr} \rho^2(x, T) dx \leq (\alpha T)^m \int_{\mathbb{R}^n} e^{cr} \rho^2(x, T_m) dx.$$

Since $\alpha T < 1$, we conclude that $\rho(x, t) \equiv 0$ for a.e $(x, t) \in \mathbb{Q}_T$. $\qquad \square$

APPENDIX B
PROOF OF THEOREM 5.4

Proof of Theorem 5.4: Let $v = \rho - \rho_R$ as in the proof of Theorem 3.1. By the maximum principle, we have that $v \geq 0$ for all $(x, t) \in B_R \times [0, T]$. Choose the test function ψ in Lemma 4.1 as

$$\Phi(x) = \gamma(x) \varrho(x),$$

where $\gamma(x) = e^{\frac{1}{2}\phi_1(x)}$ and $\phi_1(x)$, $\varrho(x)$ are defined in the proof of Proposition 5.2 and Theorem 3.1. It follows directly that $\Phi|_{\partial B_R} = \nabla_x \Phi|_{\partial B_R} = 0$, by the fact that $\varrho|_{\partial B_R} = \nabla \varrho|_{\partial B_R} = 0$. Apply Lemma 4.1 to v taking place of ρ_Ω with the test function Φ, we have

$$
\begin{aligned}
&\frac{d}{dt} \int_{B_R} \Phi v \\
&= \frac{1}{2} \int_{B_R} D_w^2 \Phi v + \int_{B_R} (f - D_w K) \cdot \Phi v + \int_{B_R} \Phi N v \\
&= \frac{1}{2} \int_{B_R} (D_w^2 \gamma \varrho + 2 D_w \gamma \cdot \nabla \varrho + \gamma D_w^2 \varrho) v \\
&\quad + \int_{B_R} (f - D_w K) \cdot (\nabla \gamma \varrho + \gamma \nabla \varrho) v + \int_{B_R} \gamma \varrho N v.
\end{aligned}
$$

All the boundary integrals vanish due to the similar arguments in Theorem 3.1. Recall that $\gamma(x) = e^{\frac{1}{2}\phi_1(x)}$ and $\varrho(x) =$

$e^{-\phi_2(x)} - e^{-R}$. Direct computations yield that

$$
\begin{aligned}
&\frac{d}{dt} \int_{B_R} \Phi v \\
&= \frac{1}{2} \int_{B_R} \left[\frac{1}{2} e^{\frac{1}{2}\phi_1} \left(D_w^2 \phi_1 + \frac{1}{2} D_w \phi_1 \cdot \nabla \phi_1 \right) \varrho \right. \\
&\qquad\qquad - e^{\frac{1}{2}\phi_1} D_w \phi_1 \cdot e^{-\phi_2} \nabla \phi_2 \\
&\qquad\qquad \left. + \gamma e^{-\phi_2} \left(D_w \phi_2 \cdot \nabla \phi_2 - D_w^2 \phi_2 \right) \right] v \\
&\quad + \int_{B_R} (f - D_w K) \cdot \left(\frac{1}{2} e^{\frac{1}{2}\phi_1} \nabla \phi_1 \varrho - \gamma e^{-\phi_2} \nabla \phi_2 \right) v \\
&\quad + \int_{B_R} \gamma \varrho N v \\
&= \int_{B_R} \Phi v \left[\frac{1}{4} \left(D_w^2 \phi_1 + \frac{1}{2} D_w \phi_1 \cdot \nabla \phi_1 \right) - \frac{1}{2} D_w \phi_1 \cdot \nabla \phi_2 \right. \\
&\qquad\qquad + \frac{1}{2} \left(D_w \phi_2 \cdot \nabla \phi_2 - D_w^2 \phi_2 \right) \\
&\qquad\qquad \left. + (f - D_w K) \cdot (\frac{1}{2} \nabla \phi_1 - \nabla \phi_2) + N \right] \\
&\quad + e^{-R} \int_{B_R} \gamma v \left[-\frac{1}{2} D_w \phi_1 \cdot \nabla \phi_2 \right. \\
&\qquad\qquad + \frac{1}{2} \left(D_w \phi_2 \cdot \nabla \phi_2 - D_w^2 \phi_2 \right) \\
&\qquad\qquad \left. - (f - D_w K) \cdot \nabla \phi_2 \right] \\
&\triangleq \int_{B_R} \Phi v [\text{VI}] + e^{-R} \int_{B_R} \gamma v [\text{VII}],
\end{aligned}
$$

By the similar estimates (5.4)-(5.6), (5.7)-(5.9), we have

$$\sup_{B_R} |\text{VI}| \leq 17n \left\| GQG^T \right\|_\infty + 5|f - D_w K| + N,$$
$$\sup_{B_R} |\text{VII}| \leq 16n \left\| GQG^T \right\|_\infty + 4|f - D_w K|.$$

Hence,

$$
\begin{aligned}
\frac{d}{dt} \int_{B_R} \Phi v &\leq C \int_{B_R} \Phi v + e^{-R} \tilde{C} \int_{B_R} e^{\phi_1} v \\
&\leq C \int_{B_R} \Phi v + e^{-R} \tilde{C} \int_{B_R} e^{\phi_1} \rho \\
&\leq C \int_{B_R} \Phi v + \tilde{C} e^{-R+Ct} \int_{B_R} e^{\phi_1} \sigma_0(x) \\
&\leq C \int_{B_R} \Phi v + \tilde{C} e^{-R+Ct} \int_{\mathbb{R}^n} e^{\phi_1} \sigma_0(x),
\end{aligned}
$$

by condition (3.5), (3.6) and (5.1). By the similar argument in the proof of Theorem 3.1, where we get the estimate of $\int_{B_R} \varrho v$, we have

$$\int_{B_R} \Phi v(x, T) \leq C e^{-R} \int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}} \sigma_0(x), \qquad \text{(B.1)}$$

where C is a generic constant, which depends on T. Recall that $\varrho(x) = e^{-R[-(|x|^2/R^2-1)^2+1]} - e^{-R}$, it implies that

$$\int_{B_R} \Phi v(x, T) \geq \frac{1}{2} e^{-\frac{7}{16}R} \int_{B_{\frac{R}{2}}} \gamma v(x, T). \qquad \text{(B.2)}$$

Combine (B.1) and (B.2), we obtain that

$$\int_{B_{\frac{R}{2}}} \gamma v(x,T) \leq Ce^{-\frac{9}{16}R} \int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}} \sigma_0(x).$$

This implies that

$$\int_{B_{\frac{R}{2}}} \gamma\rho(x,T) \leq \int_{B_{\frac{R}{2}}} \gamma\rho_R(x,T)$$
$$+ Ce^{-\frac{9}{16}R} \int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}} \sigma_0(x)$$
$$\leq C(1 + e^{-\frac{9}{16}R}) \int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}} \sigma_0(x),$$

by (5.1). Let $R \to \infty$,

$$\int_{\mathbb{R}^n} \gamma\rho(x,T) \leq C \int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}} \sigma_0(x).$$

Consider the integration outside the large ball B_R,

$$e^{\frac{1}{2}\sqrt{1+R^2}} \int_{|x|\geq R} \rho(x,T) \leq \int_{|x|\geq R} \gamma\rho(x,T)$$
$$\leq C \int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}} \sigma_0(x).$$

Therefore, we reach the conclusion that

$$\int_{|x|\geq R} \rho(x,T) \leq Ce^{-\frac{1}{2}\sqrt{1+|R|^2}} \int_{\mathbb{R}^n} e^{\sqrt{1+|x|^2}} \sigma_0(x).$$

∎

REFERENCES

[1] M. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174-188, 2002.

[2] J. S. Baras, G. L. Blankenship and W. E. Hopkins, "Existence, uniqueness and asymptotics behavior of solutions to a class of Zakai equations with unbounded coefficients," *IEEE Trans. Automat. Control*, vol. AC-28, pp. 203-214, 1983.

[3] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*, Stochastic Modelling and Applied Probability, Vol. 60, Springer, 2009.

[4] A. Bensoussan, "Some existence results for stochastic partial differential equations," in *Stochastic Partial Differential Equations and Applications*, Pitman Res. Notes Math., vol. 268, Longman Scientific and Technical, Harlow, UK, 1992, pp. 37-53.

[5] A. Bensoussan, R. Glowinski and A. Rascanu, "Approximation of the Zakai equation by the splitting up method," *SIAM J. Control Optim.*, vol. 28, pp. 1420-1431, 1990.

[6] T. E. Duncan, "Probability densities for diffusion processes with applications to nonlinear filtering theory," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1967.

[7] W. Fleming and S. Mitter, "Optimal control and nonlinear filtering for nondegenerate diffusion processes," *Stochastics*, vol. 8, pp. 63-77, 1982.

[8] A. Friedman, *Partial differential equations of parabolic type*, Prentice-Hall, Englewood Cliffs, NJ, 1964.

[9] I. Gyongy and N. Krylov, "On the splitting-up method and stochastic partial differential equation," *Ann. Probab.*, vol. 31, pp. 564-591, 2003.

[10] K. Ito, "Approxiamtion of the Zakai equation for nonlinear filtering," *SIAM J. Control Optim.*, vol. 34, pp. 620-634, 1996.

[11] R. E. Kalman, "A new approach to linear filtering and prediction problems," *ASME Trans., J. Basic Eng., ser. D.*, vol. 82, pp. 35-45, 1960.

[12] R. E. Kalman and R. S. Bucy, "New results in linear prediction and filtering theory," *ASME Trans., J. Basic Eng., ser. D.*, vol. 83, pp. 95-108, 1961.

[13] F. Le Gland, "Splitting-up approximation for SPDEs and SDEs with pplication to nonlinear filtering," Lecture Notes in Control and Inform. Sci., Vol. 176, Springer, New York, 1992, pp. 177-187.

[14] X. Luo and S. S.-T. Yau, "Hermite Spectral Method to 1D Forward Kolmogorov Equation and its Application to Nonlinear Filtering Problems," *IEEE Trans. Automat. Control*, 2013. arXiv:1301.1403 Accepted for publication

[15] N. Nagase, "Remarks on nonlinear stochastic partial differential equations: An application of the splitting-up method," *SIAM J. Control Optim.*, vol. 33, pp. 1716-1730, 1995.

[16] R. E. Mortensen, "Optimal control of continuous time stochastic systems," Ph.D. dissertation, Univ. California, Berkeley, CA, USA, 1996.

[17] E. Pardoux, "Stochastic partial differential equations and filtering of diffusion processes," *Stochastics*, vol. 3, pp. 127-167, 1979.

[18] C. Rao, "Nonlinear filtering and evolution equations: fast algorithms with applications to target tracking," Ph. D. dissertation, Univ. Southern California, Los Angeles, CA, 1998.

[19] B. L. Rozovsky, "Stochastic partial differential equations arising in nonlinear filtering problems," *Usp. Mat. Nauk.*, vol. 27, pp. 213-214, 1972.

[20] S. L. Sobolev, "Applications of functional analysis in mathematical physics," *Tran. Math. Monographs*, vol. 7, AMS, Providence, RI, 1963.

[21] S. Yau and S. S.-T. Yau, "Existence and uniqueness and decay estimates for the time dependent parabolic equation with application to Duncan-Mortensen-Zakai equation," *Asian J. Math.*, vol. 2, pp. 1079-1149, 1998.

[22] S. S.-T. Yau and S.-T. Yau, "Real time solution of nonlinear filtering problem without memory II," *SIAM J. Control Optim.*, vol. 47, no. 1, pp. 163-195, 2008.

[23] M. Zakai, "On the optimal filtering of diffusion processes," *Z. Wahrsch. Verw. Gebiete,* vol. 11, pp. 230-243, 1969.