Sparse Wiener Chaos Approximations of Nonlinear Filtering with Correlated Noise

Jian Xu and Jianxun Li

Abstract—The nonlinear filtering problem is considered for the time homogeneous diffusion model with correlated noise. A numerical approach is proposed for computing approximations of the unnormalized filtering density (UFD) and the nonlinear filtering. This approach is based on the Wiener Chaos Expansion (WCE) of the solution of Zakai equation. A Sparse truncation method of WCE is introduced to simplify calculation, and the Sparse truncation error is analyzed. Based on this truncation method, the numerical approximation scheme can substantially reduce the amount of computation and storage, or improve the precision of approximations of the UFD. When \( r=1, N=5 \) and \( n=8 \), the total number of WCE coefficients will be reduced dramatically, from 1287 to 65.

I. INTRODUCTION

The optimal nonlinear filter for a general nonlinear filtering problem is usually infinite dimensional if one resorts to moment evolution based methods [1]. Several authors derived finite-dimensional filters (FDFs) for various classes of problems [2,3,4]. However, they were considering very restrictive stochastic dynamic models [4,5] or modified cost functions [6] for minimization.

It is known [7,8,9] that, under certain regularity assumptions, the unnormalized filtering density (UFD) is the solution of the Zakai filtering equation, a stochastic parabolic equation, driven by the observation process. The exact solution of this equation can be found only in some special cases, and the development of numerical schemes for solving the Zakai equation has become an area of active research.

In some applications, like target tracking, the filter estimate must be computed in real time. Such applications require filtering algorithms with fast on line computations. Because of the large amount of calculations, many of the existing numerical schemes [10,11,12] cannot be implemented in real time when the dimension of the state process is more than three.

When the parameters of the model are known in advance, the real time computations can be simplified by separating the deterministic and stochastic components of the Zakai equation and performing the computations related to the deterministic component in advance. The separation is based on the Wiener chaos decomposition of solutions of stochastic parabolic equations. An algorithm to solve the Zakai equation using this approach for the noise uncorrelated problem was suggested in [13]. The noise correlated problem was discussed in [14,15]. But the amount of calculation is still large for real time application.

The objective of the current work is to analyze the algorithm for solving the Zakai equation of the noise correlated problem using sparse Wiener chaos approximation.

The advantage of the numerical scheme in this paper is that it can substantially reduce the number of terms in Wiener chaos expansion in the spectral approach [13,14,15], so that it can be applicable practically. Another advantage compared with the previous works [10,11,12,13] is that the presents a unified treatment of both noise-correlated and noise-uncorrelated problems with possibly degenerate diffusion in the unobserved component.

The paper is organized as follows. Section II presents the nonlinear filtering problem and its solution in a generalized form. The Wiener chaos expansion of UFD is presented in Section III and the truncation and the error of the Wiener Chaos approximation are discussed in Section IV. Numerical approach of sparse Wiener chaos approximations are presented in Section V. Conclusions are presented in Section VI. Finally, The Appendix is in Section VII.

II. THE FILTERING PROBLEM

Let \( (\Omega, \mathcal{F}, \mathbb{P}) \) be a complete probability space with independent Standard Wiener processes \( W = W(t) \) and \( V = V(t) \) of dimensions \( d_1 \) and \( r \) respectively. Let \( X_0 \) be a random variable independent of \( W \) and \( V \). In the nonlinear filtering model, the unobserved \( d \)-dimensional state process \( X = X(t) \) and the \( r \)-dimensional observation process \( Y = Y(t) \) are defined by the stochastic ordinary differential equations

\[
\begin{align*}
    dX(t) &= b(X(t))dt + \sigma(X(t))dW(t) + \rho(X(t))dV(t), \quad (1) \\
    dY(t) &= h(X(t))dt + dV(t), \\
    0 &\leq t \leq T; \quad X(0) = X_0, Y(0) = 0.
\end{align*}
\]

In the above, functions \( b = b(x); \sigma = \sigma(x); \rho = \rho(x) \) and \( h = h(x), (x \in \mathbb{R}^d) \) take values in \( \mathbb{R}^{d \times d_1}; \mathbb{R}^{d \times r}; \mathbb{R}^{d \times r} \) respectively. The underlying probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) is assumed to be fixed.

The following is assumed about the model (1)-(2):

(A1) The functions \( \sigma \) and \( \rho \) are \( C_0^\infty(\mathbb{R}^d) \), that is, bounded and three times continuously differentiable on \( \mathbb{R}^d \), so that all the derivatives are also bounded. The functions \( b \) and \( h \) are \( C_0^\infty(\mathbb{R}^d) \);
(A2) the processes $W$ and $V$ are independent;
(A3) the random vector $X_0$ is independent of both $W$ and $V$ and has density $p(x) = H^x$ for $n = 0; 1; 2; \cdots$, where $H^x$ is the Sobolev space $W_2^0(R^d)$ (see, e.g., [14]). Then, by the Sobolev embedding theorem, $p(x)$ is also in $C^0_b$ for any $n$.

Under assumption (A1)-(A3) system (1)-(2) has an unique strong solution [16, Th 5.2.5 and 5.2.9].

Let $\mathcal{F}_t^Y$ be the $\sigma$-algebra generated by $Y(s), s \leq t$. Denote
\[
Z(t) = \exp \left\{ \sum_{i=1}^r \int_0^t h_i(x(s))dY_i(s) - \frac{1}{2} \sum_{i=1}^r \int_0^t |h_i(x(s))|^2ds \right\}.
\]

Then, the measure $\tilde{P}$ defined by $d\tilde{P} = Z(t)dP$ is a probability measure on $(\Omega, \mathcal{F})$. On the reference probability space $(\Omega, \mathcal{F}; \tilde{P})$, $Y(\cdot)$ is a Brownian motion independent of $x(\cdot)$, [7,8,14].

Let $f = f(x); x \in R^d$ be a square integrable measurable scalar function. Then the filtering problem for (1) and (2) can be stated as follows: find the best mean square estimate of $f(X(t))$ given the measurements $\{y(s), 0 \leq s \leq t\}$. This estimate is called the optimal filter and will be denoted by $\hat{f}(t)$.

It is known [14] that the best mean square estimate $\hat{f}(t)$ of $f(X(t))$ given the trajectory $\{y(s), 0 \leq s \leq t\}$, can be written as
\[
\hat{f}(t) = \frac{\int f(x)u(t, x)dx}{\int u(t, x)dx},
\]
where $u = u(t, x)$ is a random field called the unnormalized filtering density (UFD). The problem of estimating $f(X(t))$ is thus reduced to the problem of computing the UFD $u$. It is also known that $u = u(t, x)$ is the solution of a certain stochastic partial differential equation
\[
\begin{align*}
\frac{du(t, x)}{dt} &= L^*u(t, x)dt + \sum_{i=1}^r \mathcal{M}^*_iu(t, x)dy_i(t), \\
u(0, x) &= p(x), \\
0 &< t \leq T, x \in R^d,
\end{align*}
\] where
\[
\begin{align*}
L^*u &= \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j}((\sigma(x)\sigma(x)^T)_{ij}u + (\rho(x)(\rho(x))^T)_{ij}u) \\
&\quad - \sum_{i=1}^d \frac{\partial}{\partial x_i}(b(x)u), \\
\mathcal{M}^*_iu &= h_i(x)u - \sum_{i=1}^d \frac{\partial}{\partial x_i}(\rho_iu(x)u),
\end{align*}
\]
where $l = 1, 2, \cdots, r$.

The exact solution of the equation (5) can be found only in a few special cases, and as a result the central part of the general nonlinear filtering problem is the numerical solution of the equation. In some applications, e.g. target tracking, the solution must be computed in real time, which puts additional restriction on the corresponding numerical scheme.

### III. CHAOS EXPANSIONS

In this section, we review the chaos expansion of the unnormalized optimal filter and its density. For more details one can refer to [14,15].

For a fixed $T > 0$, let $F = (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_0 \leq t \leq T, \mathcal{P})$ be a stochastic basis with the usual assumptions.

Consider the set of multi-indices
\[
\mathcal{J} = \{\alpha = (\alpha^k, k = 1, \cdots, r, l \geq 1), \alpha^k \in \{0, 1, 2, \cdots\}, \sum_{k,l} \alpha^k < \infty\}.
\]

The set $\mathcal{J}$ is countable, and, for every $\alpha \in \mathcal{J}$, only finitely many of $\alpha^k$ are not equal to zero. For $\alpha \in \mathcal{J}$, we write
\[
|\alpha| := \sum_{k,l} \alpha^k, \\
\alpha! := \prod_{k,l} \alpha^k!,
\]
\[
d(\alpha) := \max\{l \geq 1: \alpha^l \geq 0 \text{ for some } 1 \leq k \leq r\}.
\]

Let $H^\alpha$ be the Sobolev space $\{f : (1 + |\omega|^2)^{\alpha/2} \hat{f} \in L_2(R^d)\}$, where $\hat{f} = \hat{f}(\omega)$ is the Fourier transform of $f$, $H^0 = L_2(R^d)$ with the norm $||f||_0$. The inner product in $L_2(R^d)$ can be duality between $H^1$ and $H^{-1}$, will be denoted by $\langle \cdot, \cdot \rangle_0$. Let $\{e_i, i \geq 1\}$ be an orthonormal basis in $L_2(R^d)$, so that every $e_i$ belongs to $H^1$. Fix a positive integer number $K$. Define the matrices $A^K = (A^K_{ij}, i, j = 1, \cdots, K)$ and $B^K_l = (B^K_{lij}, i, j = 1, \cdots, K; l = 1, \cdots, r)$ by
\[
A^K = (e_i, L^*e_j)_0, \\
B^K_l = (e_i, M^*_l e_j)_0.
\]

Since $e_i \in H^1$ for all $i$, the matrices are well defined. Denote the approximation $p^K(t, x)$ of $p(t, x)$ is given by
\[
u^K(t, x) = \sum_{i=1}^K u^K_i(t)e_i(x),
\]
where the vector $u^K(t) = \{u^K_i(t), i = 1, \cdots, K\}$ is the solution of the system of stochastic ordinary differential equations
\[
du^K(t) = A^Ku^K(t)dt + \sum_{i=1}^r B^K_{li}u^K(t)dy_l(t),
\]
with the initial condition $u^K(0) = (p, e_i)_0$.

Let $0 = t_0 < t_1 < \cdots < t_M = T$ be an uniform (for simplicity) partition of the interval $[0, T]$ with step $\Delta$ and let $\{m_k(t), k \geq 1\}$ be an orthonormal basis in $L_2([0, T])$.

Define random variables
\[
\xi_{kl} = \int_{t_{i-1}}^{t_i} m_k(s - t_{i-1})dy_l(s),
\]
and then, for $\alpha \in \mathcal{J}$,
\[
\xi_\alpha = \frac{1}{\sqrt{\alpha!}} \prod_{k,l} H^\alpha_{kl}(\xi_{k,l}).
\]
where
\[ H_n(t) = (-1)^n e^{-t^2} \frac{d^n}{dt^n} e^{-t^2} \]  
(13)
is \( n \)-th Hermite polynomial. The \( N \)-th Wiener chaos is the linear
combination of all \( \varphi^{K}_{\alpha}(\xi, \cdot) \) when \( |\alpha| = k \), where
\[ d(\alpha) = b, \sum_{|\alpha|=k; d(\alpha)=b} \frac{1}{\alpha!} |\varphi^{K}_{\alpha}(\Delta, \cdot)|^2 \leq k(C_0 + \varepsilon(B))e^{C_1 \Delta^{C_k/2}} \|p\|^2 \Delta^k (b - 1)^2 (k - 1)! \]  
(22)

\[ m_1(s) = \frac{1}{\sqrt{\Delta}}, \quad m_k(s) = \sqrt{\frac{2}{\Delta}} \cos\left(\frac{\pi(k - 1)}{\Delta} s\right), \]  
k > 1, 0 ≤ s ≤ \( \Delta \).

**Definition 4.1** The filtering model (1)-(2) is called \( \nu \)-regular for some positive integer \( \nu \) if the functions \( \sigma \) and \( \rho \) belong to \( C_{2\nu+1}^{\nu+1} \), the functions \( b \) and \( h \) belong to \( C_{b+2}^{b+2} \), and \( N^\nu \) belongs to \( H^1 \), with \( \Lambda \) as in (19).

The result of S. Lototsky [14,15] is as follows:

**Lemma 4.2** (Simple Truncation). If the filtering model (1)-(2) is \( \nu \)-regular, in the sense of definition 4.1, for some \( \nu > d + 1 \)
then
\[ C_\rho = \max_{\nu, T} \sup_{x \in \mathbb{R}^d} |\rho_{\nu, T}(x)|^2, \]

the number \( C(\nu, T) \) depends on \( \nu \), \( T \) and the parameters of the model (coefficients of the equations (1)-(2)); the number \( C \) depends only on the parameters of the model.

Since the WCE decays faster in terms of \( N \) than in \( n \) and \( K \), in practice we always choose \( n > N \) and \( K > N \), so that the two terms in the error bound (21) will balance each other.

The number of the preliminary computations and the storage space are controlled by the size of \( K^2 |\mathcal{J}_N^k| = K^2 (\frac{N+n}{2})^n \). The number \( |\mathcal{J}_N^k| = \frac{(N+n)!}{N!n!} \), the size of the set \( \mathcal{J}_N^k \), increases algebraically, and very fast with respect to both \( N \) and \( n \). For example, if \( r = 1, N = 5, n = 8 \), the number of the elements in the set \( \mathcal{J}_N^k \) is \( |\mathcal{J}_5^8| = 13 \times \frac{5!8!}{2^14!} = 1287 \), and this number more than doubles for \( r = 2 \). On the other hand, if \( r = 1, N = 5, n = 8 \), then there are 1287 equations for the WCE propagator (15), which is almost impossible to solve numerically. However, we have observed that the simple truncation (18) is not optimal and we should use an adaptive truncation strategy to reduce the number of the WCE coefficients. We named it Sparse truncation. The idea of the Sparse truncation is similar to the one used by Thomas Y. Hou [17] in the numerical solutions for randomly forced equations of fluid mechanics. The following theorem is the justification for such a consideration.

**Lemma 4.3** Suppose that assumptions (A1)-(A3) hold and the deterministic basis \( m_k \) is chosen as (20), then for the WCE coefficients, we have that, when \( |\alpha| = k, d(\alpha) = b \),
\[ \sum_{|\alpha|=k; d(\alpha)=b} \left| \frac{1}{\alpha!} \varphi_{\alpha}(\Delta, \cdot) \right|^2 \leq k(C_0 + \varepsilon(B))e^{C_1 \Delta^{C_k/2}|p|^2} \frac{\Delta^k}{(b - 1)^2 (k - 1)!} \]  
(22)
Proof: Proof of this theorem will be given in the appendix.

Lemma 4.3 reveals the asymptotic decaying rate of the WCE coefficients \( u_\alpha = \varphi_\alpha / \sqrt{\alpha!} \) in terms of the polynomial order \( d(\alpha) \).

For example, let \( r = 1 \), in the truncation (18) the WCE coefficient with \( \alpha = (N, 0, ..., 0) \) has a magnitude of \( O\left(\frac{N^{N+1} \Delta N^N}{(N+1)!}\right) \), while the coefficient with \( \alpha = (0, ..., 0, N) \) has a magnitude of \( O\left(\frac{N^{N+1} \Delta N^N}{(N-1)!}\right) \).

The coefficient \( u_{(0,...,0,N)} \) is much smaller than the coefficients \( u_{(N,0,...,0)} \). Instead of using the same maximum order of Wick polynomials for every \( \xi_\alpha \), it makes more sense to use lower order Wick polynomial for \( \xi_\alpha \) with larger subscript \( |\alpha| \). For this purpose, we define an adaptive index

$$ q = (q_1, q_2, ..., q_n), \quad q_0 \leq ... \leq q_2 \leq q_1 = N. $$

In addition to the simple truncation (18), we keep only those Wick polynomials \( \xi_\alpha \) such that \( \alpha_i \leq q_i \). More precisely, we truncated the WCE expansion by the new index set

$$ J_{N,n}^q = \left\{ \alpha \in J : d(\alpha) \leq n; |\alpha| \leq N; \alpha_i \leq q_i, 1 \leq i \leq n \right\} \quad (24) $$

We choose \( q_i \) so that all the WCE coefficients have similar magnitudes. So, from Lemma 4.3 and the relevant analysis, the \( q_k \) is determined by the balance condition

$$ q_1 = N, \quad \frac{q_i}{(i-1)^2(q_i - 1)!} \approx \frac{N}{(N-1)!} \quad (25) $$

By balancing the magnitudes of the smallest coefficients, we can effectively exclude unimportant coefficients and hence reduce the dimension of the WCE truncations. To estimate the optimal \( q_i \), we just need to solve the equation (25) numerically. Since in practice we always choose \( n > N \) and \( K > N \), so that the two terms in the error bound (21) will balance each other, in real computations, an adaptive index can be set as

$$ q = (N, N, N - 1, ..., 2, 1, ..., 1), $$

or

$$ q = (N, N, N - 2, N - 4, ..., 3, 1, ..., 1). \quad (26) $$

In the above adaptive truncations, we excluded all the WCE coefficients with magnitudes less than \( O\left(\frac{N^{N+1} \Delta N^N}{(N+1)!}\right) \), which has the same order as the first term in the error estimate (21). So the adaptive truncation will not change the asymptotics of the second term in the bracket of error estimate (21). Note that the adaptive truncation deals with how to choose the optimal maximum polynomial order for each \( \xi_\alpha \), it does not change the number of retained \( \xi_\alpha \). Since the second term in the in the bracket of error estimate (21) do not depend on the polynomial order, it would not be affected by the adaptive truncation. So the adaptive truncation will not change the asymptotical behavior of the error bound (21) for the simple truncation.

Based on the above arguments, we have the following conclusion:

Theorem 4.4 (Sparse Truncation). Suppose that assumptions (A1)-(A3) hold and the deterministic basis \( m_k \) is chosen as (20). Denote the sparse Wiener chaos approximations of UFD as

$$ u_{N,n}^K(t_i) = \sum_{\alpha \in J_{N,n}^q} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha (\Delta, u_{N,n}^K(t_{i-1})) \xi_\alpha, \quad (27) $$

where \( J_{N,n}^q \) is defined by (24), and \( q = (q_1, q_2, ..., q_n) \) is determined by the balance condition (25). Then the sparse Wiener chaos approximations \( u_{N,n}^K \) has the same asymptotic accuracy as the simple truncation \( u_{N,n}^K \) defined by (18).

Using the sparse truncation, the number of the WCE coefficients can be reduced dramatically while keep the same asymptotic convergence rate. For the simple truncation \( r = 1, N = 5 \) and \( n = 8 \), the total number of WCE coefficients would be \( \frac{(N+n)!}{N!n!} = 1287 \). So the resulting WCE propagator (15) will have 1287 equations, which is almost impossible to solve numerically. But, using the sparse truncation, the number of coefficients will be reduced dramatically, from 1287 to 65. The more great of \( N \) and \( n \), the more obvious of the effect. That is a very substantially dimension reduction and the resulting WCE propagator can be solved easily.

V. SPARSE WIENER CHAOS APPROXIMATION

In this section, we will construct a recursive approximation of the solution of Zakai equation. Now, based on the spectral separating scheme [14,15] and Theorem 4.4, we can present an sparse Wiener chaos approximations algorithm for solving the Zakai equation (3) in the filtering model (1)-(2).

Let \( \{e_k\} \) be an orthonormal basis in \( L_2(\mathbb{R}^d) \) so that every function \( e_k = e_k(x) \) belongs to \( \mathbb{H}^{1} \). Define random variables \( \xi_\alpha \) according to (12).

The following is an algorithm for computing the approximations of the unnormalized filtering density and filter using Lemma 3.1 and Theorem 4.4.

1. Preliminary computations (before the observations are available):

   1. Choose suitable basis functions \( \{e_k, k = 1, \cdots, K\} \in L_2(\mathbb{R}^d) \), \( \{m_k, k = 1, \cdots, n\} \in L_2([0, \Delta]), \) and a standard unit basis \( \{U_j, j = 1, \cdots, K\} \) in \( \mathbb{R}^K \), that is, \( u_j^1 = 1, u_j^1 = 1 \), otherwise.

   2. For \( \alpha \in J_{N,n}^q \) and \( j, k = 1, \cdots, K \), compute \( z_{jk}^{K,\alpha} = \varphi_{\alpha,j}^K(\Delta; u_k) \) (using (15)),

   $$ f_k = \int_{\mathbb{R}^d} e_k(x) dx, \quad z_{jk}^{K,\alpha}(t_0) = \int_{\mathbb{R}^d} p(x) e_k(x) dx. $$

2. Real-time computations, \( i \)-th step (as the observations become available):

   Compute \( \{\xi_\alpha^i, \alpha \in J_{N,n}^q\} \) according to (12) and update the coefficients \( \varrho_{N,n}^K (\xi_i) \) as follows: for \( j = 1, \cdots, K \),

   $$ Q_{jk}^{i}(\xi_i) = \sum_{\alpha \in J_{N,n}^q} z_{jk}^{K,\alpha} \xi_\alpha, \quad (28) $$

   $$ \varrho_{N,n}^K (t_i) = \sum_{k=1}^{K} Q_{jk}^{i}(\xi_i) \varrho_{N,n}^K (t_{i-1}), \quad (29) $$

3836
then, if necessary, compute
\[ u_{K,q}^{N,n}(t_i, x) = \sum_{j=1}^{K} \varrho_{K,n}^{N,j}(t_i) e_j(x), \tag{30} \]
and/or
\[ \phi_{K,n}^{N,i}[f] = \frac{\phi_{K,n}^{N,i}[f]}{\phi_{K,n}^{N,i}[1]}, \tag{31} \]
\[ \tilde{f}_{K,n}^{N,i} = \phi_{K,n}^{N,i}[f]. \tag{32} \]

It should be noted that the difference between WCE and sparse WCE is the choice of index set \( J_N \). On the one hand, the sparse Wiener chaos approximations algorithm, presented here, can reduce the number of the WCE coefficients dramatically while keep the same asymptotic convergence rate as the simple truncation. Consequently, the amount of compute and storage will be substantially reduced. On the other hand, from theorem 4.4 and the corresponding analysis in section IV, when the number of the WCE coefficients is same, the Sparse Wiener Chaos approximations algorithm will have greater precision than the spectral separating scheme.

VI. CONCLUSION

Sparse Wiener chaos approximations of the stochastic partial differential equation about nonlinear filtering is introduced.

The truncation method of Sparse Wiener chaos is given, which substantially reduce the amount of the computations and the storage space. The superabundance of Sparse Wiener chaos approximation error is analyzed. Although the number of terms of the WCE is substantial reduced, the approximation error does not increase.

The recursive algorithm of the solution of Zakai equation is presented. The algorithm can reduce the amount of compute and storage dramatically while keep the same asymptotic convergence rate. In other words, when the number of the WCE coefficients is same, the algorithm in this paper will have greater precision than the spectral separating scheme.

VII. APPENDIX

We begin with some auxiliary constructions.

Every multi-index \( \alpha \) with \(|\alpha| = k \) can be identified with the set \( K_{\alpha} = \{(i_1, q_1), \ldots, (i_k, q_k)\} \), so that \( i_1^0 \leq i_2^0 \leq \cdots \leq i_k^0 \) if \( i_1^0 = i_2^0 = \cdots = i_k^0 \). In the future, when there is no danger of confusion, the superscript \( \alpha \) in \( i \) and \( q \) will be omitted so that \((i_j, q_j)\) will be written instead of \((i^0_j, q^0_j)\).

Let \( P^{k} \) be the permutation group of the set \( \{ 1, \ldots, k \} \). For a given \( \alpha \in \mathcal{J} \) with \(|\alpha| = k \) and the characteristic set \( K_{\alpha} = \{(i_1, q_1), \ldots, (i_k, q_k)\} \) define
\[ E_{\alpha}(s^k; t^k) := \sum_{\sigma \in P^{k}} m_{i_1}((s_{\sigma(1)})_{t_{\sigma(1)}=q_1}) m_{i_2}((s_{\sigma(2)})_{t_{\sigma(2)}=q_2}) \cdots m_{i_k}((s_{\sigma(k)})_{t_{\sigma(k)}=q_k}), \tag{33} \]

The following notations are introduced to simplify the further presentation:
\begin{itemize}
  \item \( s^k \), the ordered set \( (s_1, s_2, \ldots, s_k) \); \( d^k := ds_1 \cdots ds_k \);
  \item \( \int_{0}^{k} \cdots \int_{0}^{k} ds^k := \int_{0}^{t_1} \cdots \int_{0}^{t_k} (\cdots)ds_1 \cdots ds_k \);
  \item \( \sum_{i^k} := \sum_{i_1, \ldots, i_k = 1}^{r} \).
\end{itemize}

Note that
\[ \int_{0}^{k} ds^k = \frac{k^k}{k!}; \quad \sum_{i^k} = r^k. \]

Proof of Lemma 4.3:

If \( \alpha \) is a multi-index with \(|\alpha| = k \) and the characteristic set \( K_{\alpha} = \{(i_1, q_1), \ldots, (i_k, q_k)\} \), then \( i_k = d(\alpha) \), the order of \( \alpha \) and so the set \( J_N^{\alpha} \) can be described as \( \{ \alpha \in \mathcal{J} : |\alpha| \leq N; i_{|\alpha|} \leq n \} \).

Since \(|\alpha(i_{|\alpha|}, q_{|\alpha|})| = |\alpha| - 1 \) and \( \alpha! \geq \alpha(i_{|\alpha|}, q_{|\alpha|})! \), it now follows from the proof of proposition 7.3 in [15] that
\[ \sum_{|\alpha|=k,d(\alpha)=b} \frac{1}{|\alpha|!} |\varphi_{\alpha}(\Delta, \cdot)|^2 \]
\[ = \sum_{q_1=1}^{r} \sum_{q_2=1}^{r} \cdots \sum_{q_k=1}^{r} \frac{1}{|\alpha|!} \sum_{k=1}^{r} \int_{0}^{(k, 1)} (f_b^{(1)} + f_b^{(2)}) E_{\alpha(b, q_k)} dt |k-1|^2 \]
\[ \leq \sum_{q_1=1}^{r} \sum_{q_2=1}^{r} \cdots \sum_{q_k=1}^{r} \frac{1}{|\alpha|!} \sum_{k=1}^{r} \int_{0}^{(k-1, \Delta)} (f_b^{(1)} + f_b^{(2)}) E_{\alpha(b, q_k)} dt |k-1|^2 \]
and the proof of Proposition 7.1 in [15] shows that the last expression is equal to
\[ \sum_{q_1=1}^{r} \sum_{q_2=1}^{r} \cdots \sum_{q_k=1}^{r} \int_{0}^{(k-1, \Delta)} |f_b^{(1)} + f_b^{(2)}| E_{\alpha(b, q_k)} dt |k-1|^2 \]
where
\[ |f_b^{(1)}|^2 = 0, k = 1; \quad |f_b^{(1)}|^2 \leq C_1 \Delta \varepsilon(B)|p|^2 \frac{k^k}{(b-1)^2}, k \geq 2; \]
\[ |f_b^{(2)}|^2 \leq C_0 e^{C_1 \Delta} |p|^2 \frac{k^k C_2^3}{(b-1)^2} \]
so, since \( \int_{0}^{(k-1, \Delta)} ds = \Delta^{k-1} / (k-1)! \), yield
\[ \sum_{|\alpha|=k,d(\alpha)=b} \frac{1}{|\alpha|!} |\varphi_{\alpha}(\Delta, \cdot)|^2 \]
\[ \leq k(C_0 + \varepsilon(B)) e^{C_1 \Delta} C_2 |p|^2 \frac{\Delta^k}{(b-1)^2 (k-1)!} \]

One can refer [15] for the more information about constant \( \varepsilon(B), C_0, C_1, C_2 \).

VIII. ACKNOWLEDGMENTS

The authors gratefully acknowledge the contribution of National Research Organization and reviewers’ comments.
REFERENCES