JOURNAL OF INTEGRAL EQUATIONS
AND APPLICATIONS
Volume 23, Number 1, Spring 2011

SPARSE DISCRETIZATION MATRICES FOR VOLTERRA INTEGRAL OPERATORS WITH APPLICATIONS TO NUMERICAL DIFFERENTIATION

HAIZHANG ZHANG AND QINGHUI ZHANG

Communicated by Yuesheng Xu

ABSTRACT. We consider Volterra integral equations having a finite dimensional feature space. This provides us flexibility to construct an orthonormal basis with small support that can be preserved by the Volterra integral operator. Under the projection method, such a basis yields a sparse discretization matrix for the Volterra integral operator. When the feature space is refinable, we introduce a construction of such an orthonormal basis from existing references. Finally, we present applications to numerical differentiation for which we obtain a quasi-linear lossless compression of the discretization matrix.

1. Introduction. Denote by $\mathbf{R}$ the field of real numbers, on which we shall work throughout the paper. Let $K : [0,1] \times [0,1] \to \mathbf{R}$ be a continuous function and A the Volterra integral operator from $L^2[0,1]$ to itself defined by the kernel K as

$$(1.1) \qquad (Af)(t) := \int_0^t K(t,s)f(s)\,ds, \quad t \in [0,1],\ f \in L^2[0,1].$$

Given $g \in L^2[0,1]$, we consider the Volterra integral equation of the first kind

$$(1.2) \qquad Au = g,$$

2010 AMS *Mathematics subject classification.* Primary 45D05, Secondary 45E10, 65D25.

Keywords and phrases. Volterra integral equations, projection methods, lossless compression, numerical differentiation, refinable vector fields, wavelet bases on invariant sets.

The second author is the corresponding author.

The first author was partially supported by Guangdong Provincial Government of China through the "Computational Science Innovative Research Team" program. The second author was supported by the Natural Science Foundation of China under grant 11001282 and by the Fundamental Research Funds for the Central Universities.

Received by the editors on January 18, 2008, and in revised form on August 16, 2008.

DOI:10.1216/JIE-2011-23-1-137

where u is an unknown function in $L^2[0,1]$ to be solved. Widely used methods for solving (1.2) include the Tikhonov regularization method, regularization methods of a Volterra type, and iterated methods. For an elegant survey and extensive collection of references on these methods, see [**17**].

Among all those available methods, the Tikhonov regularization [**14, 15, 17, 26, 31**] stands out for its advantages in convergence analysis and a posteriori selections of regularization parameters. However, the discretization scheme in the Tikhonov regularization method for (1.2) usually leads to a full matrix (see, for example, [**17, 20**]). Consequently, computational complexity in the Tikhonov regularization method might be much affected. Regularization methods of a Volterra type for (1.2) that are able to yield a lower triangular discretization matrix can be found in [**17**] and the references cited therein. There is, however, rarely a complete theoretical analysis of these methods except for some particular cases where the operator A is one-smoothing [**17**] or accretive [**27**]. Even in those special cases, no a posteriori parametric selection strategies are available when the smoothness of the exact solution is unknown [**20**].

We shall follow the Tikhonov regularization method in this paper while still aiming at a sparse discretization matrix. We consider the special situation that the kernel K has a finite dimensional *feature space*

$$\mathbf{S}_K := \operatorname{span}\{K(t,\cdot) : t \in [0,1]\}. \tag{1.3}$$

This gives us flexibility to construct an orthonormal basis for $L^2[0,1]$ such that the basis functions have small support and the majority of them is orthogonal to the feature space $\mathbf{S}_K$. As will be seen by Lemma 2.1, the support of those basis functions is preserved by the operator A. It is hence expected that the discretization matrix of the Tikhonov regularization method for (1.2) is sparse. The multilevel augmentation method developed in [**8**] can then be employed to solve the linear system resulting from the Tikhonov regularized equation of (1.2). The method is based on a multiscale decomposition of the Hilbert space $L^2[0,1]$ and a matrix splitting scheme. It provides fast, accurate and stable numerical algorithms for solving the linear system [**7–10**].

An important example of (1.2) for which the kernel K has a finite dimensional feature space comes from the classical problem of numerical differentiation [**1, 13, 20, 29, 30, 33, 35**]. Suppose that g is a smooth function on the unit interval $[0,1]$ and we have its noisy data g_δ. Note that if we only know the noisy sample of g at a discrete set of points then g_δ can be obtained by an interpolation of the discrete sample data (see, for example, [**20**]). Our purpose is to approximate the derivative $g^{(\nu)}$ of g using the known data g_δ. Three kinds of methods have been proposed in the literature for this problem: difference methods [**2, 20**], interpolation methods [**28, 30**] and regularization methods [**13, 16, 20, 29, 34**] and their equivalences [**1**].

The problem of numerical differentiation can be formulated into solving a Volterra integral equation of the first kind. To see this, we recall by the Taylor series of g at the origin with a Lagrange remainder that

$$g(t) - \sum_{j=0}^{\nu-1} \frac{g^{(j)}(0)}{j!} t^j = \int_0^t g^{(\nu)}(s) \frac{(t-s)^{\nu-1}}{(\nu-1)!}\, ds, \quad t \in [0,1].$$

Suppose that the initial data $g^{(j)}(0)$, $j = 0, 1, \ldots, \nu-1$, are known or calculated in advance. Without loss of generality, let us assume that there holds

$$g^{(j)}(0) = 0, \quad j = 0, 1, \ldots, \nu-1.$$

Under the above initial condition, one can see that $g^{(\nu)}$ is the unique solution of the following Volterra integral equation of the first kind

$$g(t) = (Au)(t) := \int_0^t K_\nu(t,s) u(s)\, ds, \quad t \in [0,1], \tag{1.4}$$

where the integral kernel K_ν is defined by

$$K_\nu(t,s) = \frac{(t-s)^{\nu-1}}{(\nu-1)!}, \quad t, s \in [0,1]. \tag{1.5}$$

When $\nu = 1$, the above Volterra integral equation approach for approximating g' was discussed, for example, in [**14, 20, 26, 29**]. Our formulation here enables us to deal with derivatives of any order by the Tikhonov regularization method. Note that the feature space $\mathbf{S}_{K_\nu}$ is the space of all the polynomials of degree at most $\nu - 1$ and has dimension ν. Equation (1.4) hence falls into our special consideration.

Therefore, besides enjoying the general advantages of the Tikhonov regularization method, our approach will lead to a quasi-linear lossless compression of the discretization matrix for (1.4).

The exposition of the paper is organized as follows. We introduce the Tikhonov regularization and elaborate our motivation in the next section. Kernels K for which the feature space (1.3) is finite dimensional are also characterized in Section 2. Under the assumption that the finite dimensional feature space is refinable with respect to a class of contraction mappings on $[0,1]$, we present in Section 3 an orthonormal basis for $L^2[0,1]$ with the two desired properties mentioned above. Using this basis, we shall obtain a quasi-linear lossless compression of the discretization matrix of the Tikhonov regularization method for (1.2). In the last section, we discuss applications to numerical differentiation.

2. Tikhonov regularization and feature spaces of finite dimension. The problem of solving (1.2) is usually ill-posed [**5, 15, 17**] in the sense that either g is not in the range of A, A is not injective or A does not have a bounded inverse. We hence turn to its minimum norm least square question aiming at finding $u_* \in L^2[0,1]$ such that

$$A^*Au_* = A^*g, \quad \|u_*\| = \min\{\|u\| : A^*Au = A^*g\}, \tag{2.1}$$

where A^* denotes the adjoint operator of A and $\|u\| := (u,u)^{1/2}$ with $(\cdot,\cdot)$ being the inner product on $L^2[0,1]$ defined for all $u, v \in L^2[0,1]$ as

$$(u,v) := \int_0^1 u(t)v(t)\,dt.$$

We assume that the orthogonal projection of g into the closure of $\operatorname{ran} A$ lies in $\operatorname{ran} A$. Thus there exists a unique $u_* \in L^2[0,1]$ satisfying (2.1).

It often occurs that we only have at hand a noisy data $g_\delta \in L^2[0,1]$ of g such that $\|g_\delta - g\| \le \delta$, where δ is a positive constant representing the level of the noise. By the ill-posedness of the problem, we use the Tikhonov regularized solution $u_{\delta,\alpha}$ of

$$(\alpha I + A^*A)u_{\delta,\alpha} = A^*g_\delta \tag{2.2}$$

as an approximation of u_*. Here I is the identity operator on $L^2[0,1]$ and α is a positive regularization parameter. Typical approaches of solving the operator equation (2.2) involve the projection method. To present the method, we let $\mathbf{N}$ be the set of all the positive integers, $\mathbf{Z}_+ := \mathbf{N} \cup \{0\}$ and assume that we have a nested sequence of subspaces $V_n \subseteq L^2[0,1]$, $n \in \mathbf{Z}_+$, whose union is dense in $L^2[0,1]$. We also set P_n the orthogonal projection from $L^2[0,1]$ onto V_n, $\mathcal{A}_n := P_n A^* A P_n$ and $g_\delta^n := P_n A^* g_\delta$. The projection method is to obtain an approximation $u_{\delta,\alpha}^n \in V_n$ of $u_{\delta,\alpha}$ by solving the equation

$$(\alpha I + \mathcal{A}_n) u_{\delta,\alpha}^n = g_\delta^n. \tag{2.3}$$

The success of the above Tikhonov regularized projection method is assured by the well-known fact that

$$\lim_{\substack{\alpha+\delta\alpha^{-1/2}\to 0\\ n\to\infty}} \|u_{\delta,\alpha}^n - u_*\| = 0.$$

For simplicity, we shall enumerate finite sets with $\mathbf{N}_n := \{1, \dots, n\}$ and later on with $\mathbf{Z}_n := \{0, \dots, n-1\}$, $n \in \mathbf{N}$. Let d_n be the dimension of V_n and w_{ni}, $i \in \mathbf{N}_{d_n}$ an orthonormal basis for V_n. The solution $u_{\delta,\alpha}^n$ of (2.3) is determined by its Fourier coefficients $x_{\delta,\alpha}^{n,i} := (u_{\delta,\alpha}^n, w_{ni})$, $i \in \mathbf{N}_{d_n}$. Similarly, we define $y_\delta^{n,i} := (g_\delta^n, w_{ni})$, $i \in \mathbf{N}_{d_n}$. The vectors $x_{\delta,\alpha}^n := [x_{\delta,\alpha}^{n,i} : i \in \mathbf{N}_{d_n}]$ and $y_\delta^n := [y_\delta^{n,i} : i \in \mathbf{N}_{d_n}]$ satisfy the linear system

$$(\alpha I_{d_n} + M_n) x_{\delta,\alpha}^n = y_\delta^n, \tag{2.4}$$

where I_{d_n} denotes the $d_n \times d_n$ identity matrix and M_n is the *discretization matrix* defined as

$$M_{n,i,j} := \int_0^1 (Aw_{ni})(t)(Aw_{nj})(t)\,dt, \quad i,j \in \mathbf{N}_{d_n}. \tag{2.5}$$

A key issue in the Tikhonov regularized projection method described above is computational complexity, which depends mainly on the sparsity of M_n. It is desirable to choose the trial space V_n and its basis so that M_n does not have too many nonzero entries. As in the finite element methods [**11**], we are inclined to select basis functions that have small support. The major difficulty is that even a function $f \in L^2[0,1]$ has small support, after operation by an integral operator, the support of Af may spread all over the unit interval $[0,1]$. Consequently, one usually ends up with a full matrix (see, for example, [**17, 20**]). Our main idea is to use basis functions whose support will be preserved by the integral operator A. It is motivated by the following simple observation.

Recall that a function $f \in L^2[0,1]$ is said to be supported on a subinterval $[\beta,\gamma] \subseteq [0,1]$ if it vanishes almost everywhere on $[0,1] \setminus [\beta,\gamma]$.

Lemma 2.1. *If $f \in L^2[0,1]$ is orthogonal to $\mathbf{S}_K$ and is supported on $[\beta,\gamma] \subseteq [0,1]$ then function Af is also supported on $[\beta,\gamma]$.*

Proof. Since f vanishes almost everywhere on $[0,\beta]$, it is clear by definition (1.1) that $(Af)(t) = 0$ for $t \in [0,\beta]$. Likewise, since f equals zero almost everywhere on $[\gamma,1]$, we observe for each $t \geq \gamma$ that

$$(Af)(t) = \int_0^t K(t,s)f(s)\,ds = \int_0^\gamma K(t,s)f(s)\,ds = (K(t,\cdot),f).$$

The last term above vanishes by the assumption that f is orthogonal to $\mathbf{S}_K$. $\square$

Denote by $\operatorname{supp} f$ the support of a function $f \in L^2[0,1]$. We say that the Volterra integral operator A *preserves* the support of $f \in L^2[0,1]$ if

$$\operatorname{supp} Af \subseteq [\min \operatorname{supp} f, \max \operatorname{supp} f].$$

We see by Lemma 2.1 that A preserves the support of functions that are orthogonal to $\mathbf{S}_K$. A simple fact below justifies such a consideration.

Proposition 2.2. *There does not exist a continuous function K on $[0,1] \times [0,1]$ for which the Volterra integral operator A defined by* (1.1) *is nontrivial and preserves the support of any function $f \in L^2[0,1]$.*

Proof. Assume that there is a continuous function K on $[0,1] \times [0,1]$ such that the Volterra integral operator A defined by (1.1) preserves the support of any function $f \in L^2[0,1]$. Fix $\beta \in (0,1)$ and $t \in [\beta,1]$.

By the assumption, we have for any $f \in L^2[0,1]$ with $\operatorname{supp} f \subseteq [0,\beta]$ such that

$$\int_0^1 K(t,s)f(s)\,ds = \int_0^t K(t,s)f(s)\,ds = (Af)(t) = 0.$$

It follows by the arbitrariness of f and the continuity of K that $K(t,s) = 0$ for every $s \in [0,\beta]$. Since β and t can be arbitrarily chosen, we get that $K(t,s) = 0$ for all $0 \leq s \leq t \leq 1$. Therefore, the Volterra integral operator A is trivial. □

For a better understanding of Proposition 2.2, we consider kernels K *of a convolution type*. Those are kernels of the form

$$K(t,s) = f(t-s), \quad t,s \in [0,1], \tag{2.6}$$

where f is a continuous function on $[-1,1]$. Set $\gamma_1 := \min\{t \in [0,1] : t \in \operatorname{supp} f\}$, $\gamma_2 := \max\operatorname{supp} f$ and $g \in L^2[0,1]$. The Titchmarsh convolution theorem [**19, 32**] says that if K has the form (2.6) and $\gamma_2 + \max\operatorname{supp} g \leq 1$ then

$$\min\operatorname{supp}(Ag) = \gamma_1 + \min\operatorname{supp} g, \quad \max\operatorname{supp}(Ag) = \gamma_2 + \max\operatorname{supp} g.$$

The above equalities yield that

$$\max\operatorname{supp}(Ag) - \min\operatorname{supp}(Ag) = (\max\operatorname{supp} g - \min\operatorname{supp} g) + (\gamma_2 - \gamma_1).$$

Note that if A is nontrivial then $\gamma_2 - \gamma_1 > 0$. Therefore, if A is nontrivial then the support of Ag will be longer than the support of g by a fixed positive length.

By Lemma 2.1 and Proposition 2.2, we require a finite dimensional feature space $\mathbf{S}_K$ in order to construct an orthonormal basis ϕ_n, $n \in \mathbf{N}$ for $L^2[0,1]$ such that

$$\lim_{n\to\infty}(\max\operatorname{supp}\phi_n - \min\operatorname{supp}\phi_n) = 0 \tag{2.7}$$

and the majority of them are orthogonal to $\mathbf{S}_K$. Let us characterize K having such a property.

Theorem 2.3. *Let K be a continuous function on $[0,1]\times[0,1]$. Then the feature space $\mathbf{S}_K$ is of finite dimension if and only if there exist finitely many pairs of continuous functions φ_j, ψ_j on $[0,1]$, $j\in\mathbf{N}_n$ such that*

$$K(t,s)=\sum_{j\in\mathbf{N}_n}\varphi_j(t)\psi_j(s),\quad s,t\in[0,1]. \tag{2.8}$$

Proof. Clearly, if K is of the form (2.8) for some continuous functions φ_j, ψ_j on $[0,1]$, $j\in\mathbf{N}_n$ then the feature space $\mathbf{S}_K$ is contained by

$$\operatorname{span}\{\psi_j : j\in\mathbf{N}_n\}$$

and is hence finite dimensional. Conversely, suppose that K is non-trivial and $\mathbf{S}_K$ is of finite dimension. Let n be the dimension of $\mathbf{S}_K$. There hence exist distinct points $t_j\in[0,1]$, $j\in\mathbf{N}_n$ such that $K(t_j,\cdot)$, $j\in\mathbf{N}_n$ constitute a basis for $\mathbf{S}_K$. We set $\psi_j:=K(t_j,\cdot)$, $j\in\mathbf{N}_n$. For each $t\in[0,1]$ there exist unique constants $c_{j,t}$, $j\in\mathbf{N}_n$ such that

$$K(t,s)=\sum_{j\in\mathbf{N}_n}c_{j,t}\psi_j(s),\quad s\in[0,1].$$

Denote by φ_j the function $t\to c_{j,t}$ $j\in\mathbf{N}_n$. The function K now has the form (2.8). It remains to prove that φ_j is continuous for each $j\in\mathbf{N}_n$. To this end, we recall that any two norms on the finite dimensional space $\mathbf{S}_K$ are equivalent. This fact yields that there exists a positive constant c satisfying for all $t,t'\in[0,1]$ that

$$\sum_{j\in\mathbf{N}_n}|\varphi_j(t)-\varphi_j(t')|^2\le c\|K(t,\cdot)-K(t',\cdot)\|^2.$$

Since $\|K(t,\cdot)-K(t',\cdot)\|$ goes to zero as t' tends to t, we prove the continuity of φ_j for each $j\in\mathbf{N}_n$. The proof is complete. □

We next consider kernels K that are *positive definite* on $[0,1]$ in the sense that for all finite distinct points $t_j\in[0,1]$, $j\in\mathbf{N}_n$ the matrix

$$[K(t_j,t_k) : j,k\in\mathbf{N}_n]$$

is symmetric and positive semi-definite. Positive definite kernels are crucial to the theory of learning [**4, 12, 36**]. We are interested in positive definite kernels with a finite dimensional feature space.

Proposition 2.4. *Let $K : [0,1] \times [0,1] \to \mathbf{R}$ be continuous. Then K is a positive definite kernel on $[0,1]$ with a finite dimensional feature space $\mathbf{S}_K$ if and only if there exist finitely many continuous functions ψ_j, $j \in \mathbf{N}_n$ such that*

$$K(t,s) = \sum_{j\in\mathbf{N}_n} \psi_j(t)\psi_j(s), \quad t,s \in [0,1]. \tag{2.9}$$

Proof. Suppose that K has the form (2.9). Then by Theorem 2.3, the feature space $\mathbf{S}_K$ is finite dimensional. We also check for any finite set $\{t_j : j \in \mathbf{N}_m\} \subseteq [0,1]$ and constants $c_j \in \mathbf{R}$, $j \in \mathbf{N}_m$ that $K(t_j,t_k) = K(t_k,t_j)$, $j,k \in \mathbf{N}_m$ and

$$\begin{aligned}\sum_{j,k\in\mathbf{N}_m} c_j c_k K(t_j,t_k) &= \sum_{j,k\in\mathbf{N}_m} c_j c_k \sum_{i\in\mathbf{N}_n} \psi_i(t_j)\psi_i(t_k)\\ &= \sum_{i\in\mathbf{N}_n}\bigg(\sum_{j\in\mathbf{N}_m} c_j\psi_i(t_j)\bigg)^2 \geq 0.\end{aligned}$$

Therefore, K is positive definite. On the other hand, suppose that K is a positive definite kernel on $[0,1]$ and the feature space $\mathbf{S}_K$ is of finite dimension. We introduce the integral operator $\mathcal{K}$ on $L^2[0,1]$ by setting for each $f \in L^2[0,1]$

$$\mathcal{K}f := \int_0^1 K(\cdot,s)f(s)\,ds.$$

The Mercer theorem in the theory of positive definite kernels (see, for example, [**12**]) asserts that $\mathcal{K}$ is a positive self-adjoint compact operator on $L^2[0,1]$ and as a consequence, there exist countable eigenfunctions ϕ_j, $j \in \mathbf{N}$ with corresponding nonnegative eigenvalues λ_j, $j \in \mathbf{N}$ such that

$$K(t,s) = \sum_{j\in\mathbf{N}} \lambda_j\phi_j(t)\phi_j(s), \quad t,s \in [0,1].$$

Since K has a finite dimensional feature space, by Theorem 2.3, the range of $\mathcal{K}$ is finite dimensional. It is implied that there are at most finitely many nonzero λ_j in the equation above. Thus we see that K is of the form (2.9). □

Finally, we discuss kernels defined by a function of one variable, that is, we investigate kernels K given by

$$K(t,s) := f(at + bs + c), \quad t, s \in [0,1], \tag{2.10}$$

where a, b, c are constants with $ab \neq 0$ and f is a continuous function on $[-|a| - |b| + c, |a| + |b| + c]$. We see that when $a = 1$, $b = -1$ and $c = 0$, K is of a convolution type. Kernels of a convolution type and a finite dimensional feature space have an important application to the Bedrosian identity, [**37**].

Proposition 2.5. *Suppose that K is given by* (2.10) *through a continuous function f. Then $\mathbf{S}_K$ is finite dimensional if and only if there exist constants $\lambda_j, \gamma_j \in \mathbf{R}$ and polynomials p_j, q_j, $j \in \mathbf{N}_n$ for some $n \in \mathbf{N}$ such that*

$$f(t) = \sum_{j \in \mathbf{N}_n} e^{\lambda_j t}\Big(\cos(\gamma_j t)p_j(t) + \sin(\gamma_j t)q_j(t)\Big), \quad t \in [-|a| - |b| + c, |a| + |b| + c]. \tag{2.11}$$

Proof. It was proved in [**3**] that if f is a continuous function on the whole $\mathbf{R}$ then

$$\operatorname{span}\{f(t - \cdot) : t \in \mathbf{R}\}$$

is of finite dimension if and only if f has the form (2.11). The proof works for the general case here. □

3. Orthonormal bases with small support. We assume in this section that $\mathbf{S}_K$ is a subspace of $L^2[0,1]$ with dimension $\nu \in \mathbf{N}$ and ϕ_j, $j \in \mathbf{N}_\nu$ form an orthonormal basis for $\mathbf{S}_K$. Our purpose is to extend this basis to an orthonormal basis ϕ_n, $n \in \mathbf{N}$ for $L^2[0,1]$ such that (2.7) holds.

Let μ be a positive integer greater than 1. We introduce μ contraction mappings on $[0,1]$ by setting

$$\psi_k(t) := \frac{t+k}{\mu}, \quad t \in [0,1],\ k \in \mathbf{Z}_\mu. \tag{3.1}$$

We call $\phi := [\phi_j : j \in \mathbf{N}_\nu] : [0,1] \to \mathbf{R}^\nu$ a *refinable vector field* [**25**] with respect to the class (3.1) of contraction mappings if there exists for each $k \in \mathbf{Z}_\mu$ a $\nu \times \nu$ matrix B_k such that

$$\phi \circ \psi_k = B_k \phi. \tag{3.2}$$

Set for each $k \in \mathbf{Z}_\mu$, $I_{k,\mu} := [(k/\mu), (k+1)/\mu]$ and denote by $\chi_{I_{k,\mu}}$ its characteristic function. We introduce μ bounded linear operators T_k, $k \in \mathbf{Z}_\mu$ on $L^2[0,1]$ as

$$T_k f := \sqrt{\mu}\chi_{I_{k,\mu}} f \circ \psi_k^{-1}, \quad f \in L^2[0,1].$$

It can be verified directly that there holds

$$T_{k'}^* T_k = \delta_{k,k'} I, \quad k, k' \in \mathbf{Z}_\mu. \tag{3.3}$$

With the settings above, we shall present a construction [**23, 24, 25**] of orthonormal wavelet bases for $L^2[0,1]$ satisfying (2.7). The construction has proven useful to collocation methods for Volterra integral equations, [**6, 7**]. To this end, we set $V_0 := \mathbf{S}_K$ and notice for all $n \in \mathbf{N}$ by (3.3) that $T_k V_{n-1}$, $k \in \mathbf{Z}_\mu$ are orthogonal to each other. We hence define recursively V_n to be the orthogonal direct sum of $T_k V_{n-1}$, $k \in \mathbf{Z}_\mu$, namely,

$$V_n := \bigoplus_{k \in \mathbf{Z}_\mu} T_k V_{n-1}, \quad n \in \mathbf{N}. \tag{3.4}$$

The refinement equations (3.2) ensure that V_{n-1} is indeed a subspace of V_n, $n \in \mathbf{N}$. Moreover, the union of V_n, $n \in \mathbf{Z}_+$ is dense in $L^2[0,1]$ if there exist constants c_j, $j \in \mathbf{N}_\nu$ such that $\sum_{j \in \mathbf{N}_\nu} c_j \phi_j = 1$, [**25**].

Let W_n be the orthogonal complement of V_{n-1} in V_n, $n \in \mathbf{N}$. By (3.4), the dimension d_n of V_n is $\nu\mu^n$. Thus, the dimension d'_n of W_n is given by

$$d'_n = d_n - d_{n-1} = \nu(\mu-1)\mu^{n-1}, \quad n \in \mathbf{N}.$$

A recursive construction of an orthonormal wavelet basis ω_{ni}, $i \in \mathbf{N}_{d'_n}$ for W_n was provided in [**23, 24, 25**]. The construction starts with an arbitrary orthonormal basis ω_{1i}, $i \in \mathbf{N}_{d'_1}$ for W_1. It follows by (3.4) that there holds

$$W_{n+1} = \bigoplus_{k \in \mathbf{Z}_\mu} T_k W_n, \quad n \in \mathbf{N}.$$

Therefore,

$$T_k \omega_{ni}, \quad i \in \mathbf{N}_{d'_n}, \quad k \in \mathbf{Z}_\mu \tag{3.5}$$

form an orthonormal basis for W_{n+1}. We reindex this basis as $\omega_{(n+1)i}$, $i \in \mathbf{N}_{d'_{n+1}}$. Under this construction, each ω_{ni}, $n \in \mathbf{N}$, $i \in \mathbf{N}_{d'_n}$ is supported on $[\bar{i}/(\mu^{n-1}), (\bar{i}+1)/(\mu^{n-1})]$, where

$$\bar{i} := (i-1) \bmod \mu^{n-1}, \quad i \in \mathbf{N}_{d'_n}.$$

Note that

$$V_n = V_0 \bigoplus_{i \in \mathbf{N}_n} W_i. \tag{3.6}$$

Thus, if we set $\omega_{0j} := \phi_j$, $j \in \mathbf{N}_\nu$ and $d'_0 := \nu$ then ω_{ij}, $(i,j) \in \mathbf{I}_n := \{(i,j) : i \in \mathbf{Z}_{n+1}, j \in \mathbf{N}_{d^p_i}\}$ form an orthonormal basis for V_n, $n \in \mathbf{N}$. Using this basis we shall get a quasi-linear lossless compression of the discretization matrix (2.5).

Theorem 3.1. *Let the trial space V_n and its basis be defined as above. Then the number of nonzero entries in the discretization matrix M_n defined by (2.5) is of $O(d_n \log d_n) = O(n\mu^n)$.*

Proof. By definition, an entry $M_{ij,i'j'}$, $(i,j), (i',j') \in \mathbf{I}_n$ has the form

$$M_{ij,i'j'} = \int_0^1 (A\omega_{ij})(t)(A\omega_{i'j'})(t)\, dt.$$

By Lemma 2.1, the support of $A\omega_{ij}$ and $A\omega_{i'j'}$ is contained in $[(\bar{j}/\mu^{i-1}), (\bar{j}+1)/(\mu^{i-1})]$ and $[\overline{j'}/(\mu^{i'-1}), (\overline{j'}+1)/(\mu^{i'-1})]$, respectively.

Therefore, $M_{ij,i'j'}$ does not vanish only if these two intervals overlap. By this observation, we calculate that the number of nonzero entries in M_n is less than or equal to

$$\nu^2(2\mu^n - 1) + \sum_{i=1}^{n} d_i' \left[\nu(\mu - 1) + 2 \sum_{i'=i+1}^{n} \nu(\mu - 1)\mu^{i'-i} \right].$$

The above quantity can be simplified as

$$2\nu^2(\mu - 1)n\mu^n - \nu^2(\mu - 1)\mu^n + \mu\nu^2. \tag{3.7}$$

Thus the number of nonzero entries in M_n is of $O(n\mu^n)$. □

With the multiscale decomposition (3.6), the linear system (2.4) or equivalently the operator equation (2.3) can be solved by the multilevel augmentation method developed in [**8**]. For a brief description of the method, see [**10**]. It starts with an exactly computed $u^k_{\delta,\alpha}$ for a chosen $k \in \mathbf{N}$ much smaller than n. The augmentation algorithm then generates an approximate solution $u^{k,n-k}_{\delta,\alpha}$ of $u^n_{\delta,\alpha}$. The algorithm has been proven to be efficient when a sparse discretization matrix M_n is available. In particular, it has been shown in [**8**] that the number of multiplications required to obtain $u^{k,n-k}_{\delta,\alpha}$ is of $O(n\mu^n)$. For the error analysis, it can be seen from the proof of Theorem 3.4 in [**10**] under some assumptions (see hypotheses H-1, H-2 therein) that there exists a constant C independent of α, δ such that

$$\|u^{k,n-k}_{\delta,\alpha} - u_*\| \le C\alpha^p + \frac{\delta}{\sqrt{\alpha}}, \tag{3.8}$$

where $p \in (0, 1]$ is a constant such that u_* is contained in the range of $(A^*A)^p$. The constant p represents the smoothness of the exact solution u_*. When this smoothness is known, a priori choice $\alpha = O(\delta^{2/(2p+1)})$ gives the optimal convergence rate $O(\delta^{2p/(2p+1)})$. When p is unknown, based on the bias-variance structure of the approximation error (3.8), we can still use the Lepskii principle for a posteriori choice the regularization parameter α to achieve the optimal convergence rate. The Lepskii principle has been extensively developed to solve ill-posed operator equations in [**18, 21, 22**] and has recently been applied to numerical differentiation in [**20**]. We refer the readers to [**20, 21, 22**] and the references cited therein for a detailed description and analysis of the principle.

We conclude that a complete numerical algorithm for the Volterra equation (1.2) has been proposed under the assumptions that the feature space $\mathbf{S}_K$ is finite dimensional and its basis functions constitute a refinable vector field. It consists of applying the Tikhonov regularization to (1.2), constructing a basis for $L^2[0,1]$ based on the multiscale decomposition (3.6), solving the linear system (2.4) by the multilevel augmentation method, and the Lepskii principle for a posteriori choice of the regularization parameter.

4. Applications to numerical differentiation. In this section, we discuss applications to numerical differentiation. Specifically, given a noisy data g_δ of g such that $\|g_\delta - g\| \leq \delta$, we shall employ the algorithm described in the last section to solve $u = g^{(v)}$ from (1.4).

Recall that for the kernel K_v given by (1.5), the feature space $\mathbf{S}_{K_v}$ is the space of all the polynomials of degree at most $\nu-1$. An orthonormal basis for $\mathbf{S}_{K_v}$ is provided by

$$\phi_j(t) := \sqrt{2j-1}P_{j-1}(2t-1), \quad t \in [0,1],\ j \in \mathbf{N}_\nu,$$

where P_n denotes the Legendre polynomials defined as

$$P_0(t) := 1, \qquad P_n(t) := \frac{1}{2^n n!}\frac{d^n}{dt^n}[(t^2-1)^n], \quad t \in \mathbf{R},\ n \in \mathbf{N}.$$

The vector field $\phi := [\phi_j : j \in \mathbf{N}_\nu]$ is refinable with respect to the contraction mappings (3.1) since for all $j \in \mathbf{N}_\nu$, $k \in \mathbf{Z}_\mu$, $\phi_j \circ \psi_k$ is still a polynomial of degree at most $\nu - 1$. The construction method in Section 3 can hence be applied to generate an orthonormal wavelet basis for $L^2[0,1]$. By Theorem 3.1, this basis results in a quasi-linear lossless compression of the discretization matrix M_n. The multilevel augmentation method and the Lepskii principle can then be employed to solve the linear system (2.4) and choose the regularization parameter, respectively.

We next present numerical experiments for the following chosen functions

$$g_1(t) := t - \frac{t^2}{2}, \quad t \in [0,1], \qquad g_2(t) := \begin{cases} -t & t \in [0,(1/2)], \\ t & t \in ((1/2),1], \end{cases}$$

$$g_3(t) := \frac{t}{3}(t-1)^3 - \frac{t^4}{4} + \frac{2}{3}t^3 - \frac{t^2}{2} + \frac{t}{3}, \quad t \in [0,1].$$

We shall calculate the first derivative of g_1, g_2 and the second derivative of g_3 with the noise level $\delta = 0.001$. To this end, we specify $\nu = 2$, our trial space to be V_9 and $\mu = 2$. The orthonormal wavelet basis for W_0 and W_1 is given respectively as

$$\omega_{01}(t) := 1, \qquad \omega_{02}(t) := 2\sqrt{3}\left(t - \frac{1}{2}\right), \quad t \in [0,1],$$

and

$$\omega_{11}(t) := -6\left(t - \frac{1}{2}\right) + 2\operatorname{sgn}\left(t - \frac{1}{2}\right),$$
$$\omega_{12}(t) := 4\sqrt{3}\left|t - \frac{1}{2}\right| - \sqrt{3}, \quad t \in [0,1],$$

where the signum function $\operatorname{sgn}(t)$ takes the value -1, 0 or 1 for $t < 0$, $t = 0$ or $t > 0$, respectively. Orthonormal wavelet bases for W_n, $n \in \mathbf{N}_9 \setminus \{1\}$ can be generated recursively from ω_{11} and ω_{12} by (3.5).

Under the above settings, the discretization matrix M_9 is of size 1024×1024. It is counted that the number of nonzero entries in M_9 is 34824, which is exactly equal to (3.7) for our choice $\mu = \nu = 2$ and $n = 9$. Thus the compression rate is approximately equal to 30 and the space savings is

$$1 - \frac{34824}{1024^2} \approx 96.68\%.$$

We sketch in Figure 1 the distribution of nonzero entries in the matrix M_9. It can be seen that they are rather sparse in M_9. In the calculation of the numerical derivatives of g_j, $j \in \mathbf{N}_3$, we use the multilevel augmentation method in [**10**] and the adaptive a posteriori choice of regularization parameters in [**18, 21**]. The results obtained are shown in Figures 2, 3 and 4, respectively. In particular, function g_2 was used in [**20, 34**] to locate the discontinuity points of derivatives. We observe from Figure 3 that our result correctly identifies the discontinuity point $t = 0.5$.

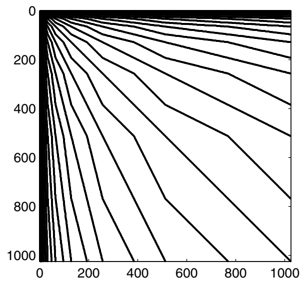

FIGURE 1. Nonzero entries in M_9.

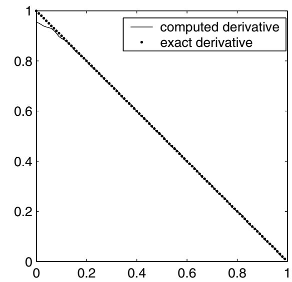

FIGURE 2. Numerical example for g_1.

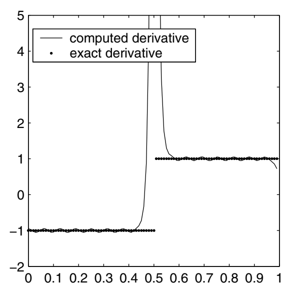

FIGURE 3. Numerical example for g_2.

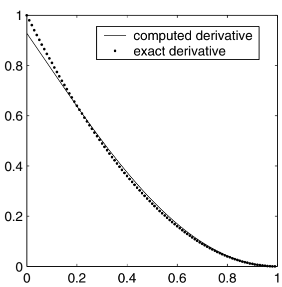

FIGURE 4. Numerical example for g_3.

For a more detailed illustration of our method, we shall choose $\nu = 1$,

$$
\begin{aligned}
g(t) &:= \sin\left(\frac{\pi}{2}t\right), \\
g_\delta(t) &:= \sin\left(\frac{\pi}{2}t\right) + \delta\cos(1000t), \quad t \in [0,1],
\end{aligned}
\tag{4.1}
$$

where various values of δ will be specified in the numerical experiments. One can see that $u_* = g'$ is contained by the range of A^*A. Thus the optimal convergence rate is $O(\delta^{2/3})$, which will be attained by our method. The regularization parameter α will be chosen by the Lepskii principle. We tabulate the numerical results below.

TABLE 1. Numerical results for $g = \sin((\pi/2)t)$.

δ	k	2^{n+1}	r	α	e	$e/\delta^{2/3}$
0.05	4	128	5.81	0.2479	0.4214	3.105
0.01	4	256	9.84	0.06527	0.1540	3.319
0.005	5	512	17.06	0.04053	0.1010	3.453
0.001	5	1024	30.12	0.01291	0.03429	3.429
0.0005	6	2048	53.76	0.008018	0.02155	3.420
0.0001	6	4096	97.09	0.002555	0.006958	3.229

The parameter k in the above table is the initial level in the multilevel augmentation method. Columns 3 and 4 give the size and compression rate of the discretization matrix, respectively. We also let e denote the error $\|u_* - u_{\delta,\alpha}^{k,n-k}\|$. The last column indicates that the optimal convergence rate $O(\delta^{2/3})$ is attained.

Finally, we compare our method with the one in the recent reference [**20**], where piecewise linear functions are used to discretize the operator equation (2.3). Since discretization by piecewise linear functions leads to a full positive definite matrix, the resulting linear system will be solved by a Cholesky decomposition of the discretization matrix, followed by solving two linear systems with a lower triangular and upper triangular coefficient matrix. The number of multiplications required is of $O(N^3)$ if the discretization matrix is of size $N \times N$. By contrast our method requires only $O(N \log N)$ multiplications.

For numerical comparison, again we shall employ example (4.1). The difference is that the regularization parameter α will be chosen as $\delta^{2/3}$ based on a priori information that g' is contained by the range of A^*A. In our numerical experiments, we fix the initial level k to be 6 and n to be 10. To assure the same convergence rate, the number of knots for the piecewise linear functions in [**20**] is chosen to be 1024. The numerical results are presented in Table 2.

TABLE 2. Numerical results for the comparison.

δ	α	e_1	t_1	e_2	t_2
0.01	0.04642	0.1141	0.086	0.1147	0.24
0.001	0.01000	0.02674	0.079	0.02743	0.25
0.0001	0.002154	0.005873	0.078	0.006635	0.25

Here, e_1 and t_1 denote the error measured by the norm of $L^2[0,1]$ and consuming time in seconds of solving the discretization linear system in our method. Parameters e_2, t_2 denote those of the method in [**20**]. As can be seen from Table 2, our method also achieves the optimal convergence rate and is superior in computational complexity. Note that the advantage in computation complexity of our method becomes more important if a posteriori choices of regularization parameters are engaged. The reason is that one usually has to select the parameter from a prescribed set. Thus several discretization linear systems need to be solved.

REFERENCES

1. R.S. Anderssen and P. Bloomfield, *Numerical differentiation procedures for non-exact data*, Numer. Math. **22** (1974), 157–182.

2. R.S. Anderssen and F.R. de Hoog, *Finite difference methods for the numerical differentiation of nonexact data*, Computing **33** (1984), 259–267.

3. P.M. Anselone and J. Korevaar, *Translation invariant subspaces of finite dimension*, Proc. Amer. Math. Soc. **15** (1964), 747–752.

4. N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc. **68** (1950), 337–404.

5. H. Brunner and P.J. van der Houwen, *The numerical solution of Volterra equations*, North-Holland, Amsterdam, 1986.

6. Y. Cao, T. Herdman and Y. Xu, *A hybrid collocation method for Volterra integral equations with weakly singular kernels*, SIAM J. Numer. Anal. **41** (2003), 364–381.

7. Z. Chen, C.A. Micchelli and Y. Xu, *Fast collocation methods for second kind integral equations*, SIAM J. Numer. Anal. **40** (2002), 344–375.

8. Z. Chen, B. Wu and Y. Xu, *Multilevel augmentation methods for solving operator equations*, Numer. Math. J. Chinese Univ. **14** (2005), 31–55.

9. ———, *Multilevel augmentation methods for differential equations*, Adv. Comput. Math. **24** (2006), 213–238.

10. Z. Chen, Y. Xu and H. Yang, *A multilevel augmentation method for solving ill-posed operator equations*, Inverse Problems **22** (2006), 155–174.

11. P.G. Ciarlet, *The finite element method for elliptic problems*, Classics Appl. Math. **40**, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002.

12. F. Cucker and S. Smale, *On the mathematical foundations of learning*, Bull. Amer. Math. Soc. **39** (2002), 1–49.

13. J. Cullum, *Numerical differentiation and regularization*, SIAM J. Numer. Anal. **8** (1971), 254–265.

14. H.W. Engl, M. Hanke and A. Neubauer, *Regularization of inverse problems*, Kluwer, Dordrecht, 1996.

15. C.W. Groetsch, *The theory of Tikhonov regularization for Fredholm equations of the first kind*, Research Notes Math. **105**, Pitman, Boston, MA, 1984.

16. M. Hanke and O. Scherzer, *Inverse problems light*: *Numerical differentiation*, Amer. Math. Monthly **108** (2001), 512–521.

17. P.K. Lamm, *A survey of regularization methods for first-kind Volterra equations*, in *Surveys on solution methods for inverse problems*, Springer, Vienna, 2000, 53–82.

18. O.V. Lepskii, *On a problem of adaptive estimation in Gaussian white noise*, Theory Prob. Appl. **35** (1990), 454–466.

19. J.L. Lions, *Support dans la transformation de Laplace*, J. Anal. Math. **2** (1952/53), 369–380.

20. S. Lu and S.V. Pereverzev, *Numerical differentiation from a viewpoint of regularization theory*, Math. Comp. **75** (2006), 1853–1870.

21. P. Mathe and S.V. Pereverzev, *Geometry of linear ill-posed problems in variable Hilbert scales*, Inverse Problems **19** (2003), 789–803.

22. ———, *Discretization strategy for linear ill-posed problems in variable Hilbert scales*, Inverse Problems **19** (2003), 1263–1277.

23. C.A. Micchelli and Y. Xu, *Using the matrix refinement equation for the construction of wavelets on invariant sets*, Appl. Comput. Harmon. Anal. **1** (1994), 391–401.

24. ———, *Using the matrix refinement equation for the construction of wavelets.* II. *Smooth wavelets on* $[0,1]$, in *Approximation and computation*, R. Zahar, ed., Birkhauser, Boston, 1995.

25. ———, *Reconstruction and decomposition algorithms for biorthogonal wavelets*, Multidimens. Syst. Signal Process. **8** (1997), 31–69.

26. A. Neumaier, *Solving ill-conditioned and singular linear systems*: *A tutorial on regularization*, SIAM Rev. **40** (1998), 636–666.

27. R. Plato, *The Galerkin scheme for Lavrentiev's m-times iterated method to solve linear accretive Volterra integral equations of the first kind*, BIT **37** (1997), 404–423.

28. R. Qu, *A new approach to numerical differentiation and integration*, Math. Comput. Modelling **24** (1996), 55–68.

29. A.G. Ramm and A.B. Smirnova, *On stable numerical differentiation*, Math. Comp. **70** (2001), 1131–1153.

30. T.J. Rivlin, *Optimally stable Lagrangian numerical differentiation*, SIAM J. Numer. Anal. **12** (1975), 712–725.

31. W.W. Schmaedeke, *Approximate solutions for Volterra integral equations of the first kind*, J. Math. Anal. Appl. **23** (1968), 604–613.

32. E.C. Titchmarsh, *The zeros of certain integral functions*, Proc. London Math. Soc. **s2-25** (1926), 283–302.

33. G. Wahba, *Cross-validated spline methods for the estimation of multivariate functions from data on functionals*, in *Statistics*: *An appraisal*, H.A. David and H.T. David, eds., Iowa State University, 1984.

34. Y.B. Wang, X.Z. Jia and J. Cheng, *A numerical differentiation method and its application to reconstruction of discontinuity*, Inverse Problems **18** (2002), 1461–1476.

35. H.J. Woltring, *A Fortran package for generalized cross-validatory spline smoothing and differentation*, Adv. Engineering Software **8** (1986), 104–113.

36. Y. Xu and H. Zhang, *Refinable kernels*, J. Mach. Learn. Res. **8** (2007), 2083–2120.

37. B. Yu and H. Zhang, *The Bedrosian identity and homogeneous semi-convolution equations*, J. Integral Equations Appl. **20** (2008), 527–568.

School of Mathematics and Computational Science, and Guangdong Province Key Laboratory of Computational Science, Sun Yat-sen University, Guangzhou 510275, P.R. China
Email address: zhhaizh2@sysu.edu.cn

School of Mathematics and Computational Science, Sun Yat-sen University, Guangzhou 510275, P.R. China
Email address: zhangqh6@mail.sysu.edu.cn