Wavelet-Galerkin Method for Integro-Differential Equations

Dana Černá, Václav Finěk

Technical University of Liberec

September 2019

Talks

1. Discrete wavelet transform, wavelets, and wavelet basis
2. Construction of spline wavelet basis with short support
3. Wavelet methods for integro-differential equations
4. Wavelet methods for option pricing

Outline

Introduction - PIDE, existence, uniqueness

Construction of quadratic spline wavelet basis

Wavelet-Galerkin method

Compression of discretization matrices

Numerical examples

Pdf presentation is at https://kmd.fp.tul.cz/en/cb-profile/cerna

The details can be found in

D. Černá, V. Finěk: Galerkin method with new quadratic spline wavelets for integral and integro-differential equations. *Journal of Computational and Applied Mathematics* **363**, 2020, pp. 426-443.

Introduction

- We focus on the numerical solution of Fredholm linear integral equations and second-order integro-differential equations using the wavelet-Galerkin method.
- We construct appropriate wavelet basis (on the interval and hyperrectangle) for this problem (satisfying required properties such as Riesz basis property, smoothness, vanishing moments, boundary conditions).
- We present (sketch of) the proof of the Riesz basis property.
- We show that discretization matrices have uniformly bounded condition numbers and that they can be approximated by sparse matrices.
- We provide numerical examples and compare the results with the Galerkin method using other wavelet bases and other methods.

Advantages of wavelet methods

- Discretization matrices can be approximated by sparse matrices, while most of the standard methods (FD, FEM, Galerkin with B-splines, collocation method, quadrature method) lead to full matrices.
- The condition numbers of the discretization matrices are uniformly bounded. This implies that the number of iterations needed to resolve a discrete problem with a desired accuracy is uniformly bounded.
- The convergence of the wavelet-Galerkin method is of high order if high-order spline wavelets are used. Order of convergence in the L^2-norm for quadratic spline wavelet basis is $\mathcal{O}\left(h^3\right)$, where h is the step of the method.
- The wavelet-Galerkin method with the constructed quadratic-spline wavelet basis was more efficient than this method with other quadratic-spline bases.

Notation

Let $\Omega = (a_1, b_1) \times (a_2, b_2) \times \ldots \times (a_d, b_d)$, i.e. Ω is a hyperrectangle.

Let $L^2(\Omega)$ be the space of real-valued square-integrable functions on Ω equipped with

$$\langle f, g \rangle = \int_\Omega f(x)\, g(x)\, dx, \quad \|f\| = \sqrt{\langle f, f \rangle}.$$

Let $H^1(\Omega)$ be the Sobolev space, i.e. the space of all functions from $L^2(\Omega)$ for which their first-order weak derivatives also belong to $L^2(\Omega)$, which is equipped with

$$\langle f, g \rangle_{H^1} = \sum_{i=1}^{d} \left\langle \frac{\partial f}{\partial x_i}, \frac{\partial g}{\partial x_i} \right\rangle + \langle f, g \rangle,$$

and

$$\|f\|_{H^1} = \sqrt{\langle f, f \rangle_{H^1}}, \quad |f|_{H^1} = \sqrt{\sum_{i=1}^{d} \left\| \frac{\partial f}{\partial x_i} \right\|^2}.$$

Let H_0^1 be the closure in $H^1(\Omega)$ of the set of all functions f such that $\operatorname{supp} f \subset \Omega$, f is continuous on $\overline{\Omega}$ and f has continuous first order derivatives in Ω. The set of all m-times continuously differentiable functions on $\overline{\Omega}$ is denoted as $C^m(\overline{\Omega})$.

Let $K \in L^2(\Omega \times \Omega)$ and let $\mathcal{K} : L^2(\Omega) \to L^2(\Omega)$ be an integral operator given by

$$(\mathcal{K}y)(t) = \int_\Omega K(t,x)\, y(x)\, dx.$$

We denote $\Delta y = \frac{\partial^2 y}{\partial x_1^2} + \ldots + \frac{\partial^2 y}{\partial x_d^2}$.

Our aim is to find a solution y of the equation

$$\mathcal{A}y := -\epsilon \Delta y + py + \mathcal{K}y = f \quad \text{on } \Omega,$$

where $\epsilon \geq 0$ is a constant.

Two cases

1. We assume that $\epsilon = 0$, i.e. our aim is to find a solution $y \in L^2(\Omega)$ of the linear integral equation

$$p(t)\, y(t) + \int_\Omega K(t,x)\, y(x)\, dx = f(t), \quad t \in \Omega.$$

2. We assume that $\epsilon > 0$ and our aim is to find a solution y of the second-order integro-differential equation

$$-\epsilon \Delta y(t) + p(t)\, y(t) + \int_\Omega K(t,x)\, y(x)\, dx = f(t), \quad t \in \Omega,$$

satisfying homogeneous Dirichlet boundary conditions $y = 0$ on the boundary of Ω.

Variational formulation

Let $V = L^2(\Omega)$ for $\epsilon = 0$ and $V = H_0^1(\Omega)$ for $\epsilon > 0$, let $\|\cdot\|_V$ be a norm in V, and let us define the bilinear form $a : V \times V \to \mathbb{R}$ as

$$a(u, v) = \epsilon \sum_{i=1}^{d} \left\langle \frac{\partial u}{\partial t_i}, \frac{\partial v}{\partial t_i} \right\rangle + \langle pu, v \rangle + \langle \mathcal{K}u, v \rangle$$

for $u, v \in V$.

The variational formulation of the equation reads as: Find $y \in V$ such that

$$a(y, v) = \langle f, v \rangle \quad \text{for all} \quad v \in V. \tag{1}$$

Assumptions

(A1) The function p satisfies $p \in C\left(\overline{\Omega}\right)$, and there exists a constant p_{min} such that $p(t) \geq p_{min} > 0$ for all $t \in \overline{\Omega}$.

(A2) The kernel K is smooth enough, i.e. $K \in C^m\left(\overline{\Omega} \times \overline{\Omega}\right)$ for some $m \in \mathbb{N}$.

(A3) The function f belongs to the space $L^2(\Omega)$.

(A4) The bilinear form a is coercive, which means that there exists a constant $\alpha > 0$ such that

$$a(u,u) \geq \alpha \|u\|_V^2 \quad \text{for all} \quad u \in V.$$

Existence and uniqueness

Theorem

If the assumptions (A1)–(A4) are satisfied, then there exists a unique solution $y \in V$ of Equation (1).

Proof.

Since

$$\left| \int_\Omega u(x)\, dx \right| \leq C \|u\|, \quad C = \sqrt{\int_\Omega 1 dx}, \tag{2}$$

we have

$$\begin{aligned} |a(u,v)| &\leq \epsilon |u|_{H^1} |v|_{H^1} + p_{max} \|u\| \|v\| + K_{max} C^2 \|u\| \|v\| \\ &\leq \max(\epsilon, p_{max} + K_{max} C^2) \|u\|_V \|v\|_V , \end{aligned}$$

where $p_{max} = \max_{x \in \overline{\Omega}} p(x)$ and $K_{max} = \max_{x,t \in \overline{\Omega}} |K(x,t)|$. The existence and uniqueness of the solution follows from the continuity and coercivity of the bilinear form a by the Lax-Milgram lemma. □

Lemma

If (A1) and (A2) are satisfied and there exists a constant K_{min} such that $K(x,t) \geq K_{min}$ for all $x, t \in \overline{\Omega}$ and $K_{min} + p_{min} > 0$, then a is coercive.

Proof.

If $u \in L^2(\Omega)$ and the assumptions of the lemma are satisfied, then

$$\langle \mathcal{K}u, u \rangle = \int_\Omega \int_\Omega K(x,t)\, u(x)\, u(t)\, dxdt \geq K_{min} \left(\int_\Omega u(x)\, dx \right)^2 .$$

If K_{min} is positive then $\langle \mathcal{K}u, u \rangle \geq 0$. Due to (2), we have $\langle \mathcal{K}u, u \rangle \geq K_{min} C^2 \|u\|^2$ for K_{min} negative. If $\epsilon = 0$, then

$$a(u,u) \geq \min\left(p_{min}, p_{min} + K_{min} C^2\right) \|u\|^2 .$$

If $\epsilon > 0$, then

$$\begin{aligned} a(u,u) &\geq \epsilon |u|_{H^1}^2 + \min\left(p_{min}, p_{min} + K_{min} C^2\right) \|u\|^2 \\ &\geq \min\left(\epsilon, p_{min}, p_{min} + K_{min} C^2\right) \|u\|_{H^1}^2 . \end{aligned}$$

Wavelet basis

Let H be a Sobolev space or the L^2–space, $\mathcal{J}$ be an index set and let $\lambda \in \mathcal{J}$ take the form $\lambda = (j,k)$. A wavelet basis of H is defined as a family $\Psi = \{\psi_\lambda, \lambda \in \mathcal{J}\}$ such that

i) Ψ is a Riesz basis for H, i.e. the closure of the span of Ψ is H and there exist constants $c, C \in (0,\infty)$ such that

$$c\,\|\mathbf{b}\|_2 \leq \left\| \sum_{\lambda\in\mathcal{J}} b_\lambda \psi_\lambda \right\|_H \leq C\,\|\mathbf{b}\|_2\,,$$

for all $\mathbf{b} = \{b_\lambda\}_{\lambda\in\mathcal{J}}$ such that $\sum_{\lambda\in\mathcal{J}} b_\lambda^2 < \infty$.

The number $\inf C / \sup c$ is called the condition number of Ψ.

ii) The functions are local in the sense that $\operatorname{diam} \mathit{supp}\ \psi_\lambda \leq C 2^{-|\lambda|}$ for all $\lambda \in \mathcal{J}$, and at a given level $j = |\lambda|$ the supports of only finitely many wavelets overlap at any point x.

Structure of the wavelet basis

A wavelet basis on the interval I has typically the hierarchical structure:

$$\Psi^I = \Phi^I_{j_0} \cup \bigcup_{j=j_0}^{\infty} \Psi^I_j.$$

$\Phi^I_{j_0} = \left\{ \phi^I_{j_0,k}, k \in \mathcal{I}_{j_0} \right\}$ - the set of scaling functions

$\Psi^I_j = \left\{ \psi^I_{j,k}, k \in \mathcal{J}_j \right\}$ - the set of wavelets

Wavelets and scaling functions in the inner part of the interval are typically translations and dilations of one or several functions. Wavelets and scaling functions near the boundary are dilations of some special functions called boundary scaling functions and wavelets.

We assume that wavelets have vanishing moments, i.e.

$$\int_I x^m \psi^I_{j,k}(x)\,\mathrm{d}x = 0, \quad m = 0, \ldots L-1, \quad k \in \mathcal{J}_j,$$

where $L \geq 1$ is dependent on the type of a wavelet.

Construction of wavelet basis

We define a scaling basis as a basis of quadratic B-splines. Let ϕ be a quadratic B-spline defined on knots $[0, 1, 2, 3]$. It can be written explicitly as

$$\phi(x) = \begin{cases} \frac{x^2}{2}, & x \in [0, 1], \\ -x^2 + 3x - \frac{3}{2}, & x \in [1, 2], \\ \frac{x^2}{2} - 3x + \frac{9}{2}, & x \in [2, 3], \\ 0, & \text{otherwise.} \end{cases}$$

Let ϕ_{b1} be a quadratic B-spline defined on knots $[0, 0, 0, 1]$ and let ϕ_{b2} be a quadratic B-spline defined on knots $[0, 0, 1, 2]$, i.e.

$$\phi_{b1}(x) = \begin{cases} x^2 - 2x + 1, & x \in [0, 1], \\ 0, & \text{otherwise,} \end{cases} \quad \phi_{b2}(x) = \begin{cases} -\frac{3x^2}{2} + 2x, & x \in [0, 1], \\ \frac{x^2}{2} - 2x + 2, & x \in [1, 2], \\ 0, & \text{otherwise.} \end{cases}$$

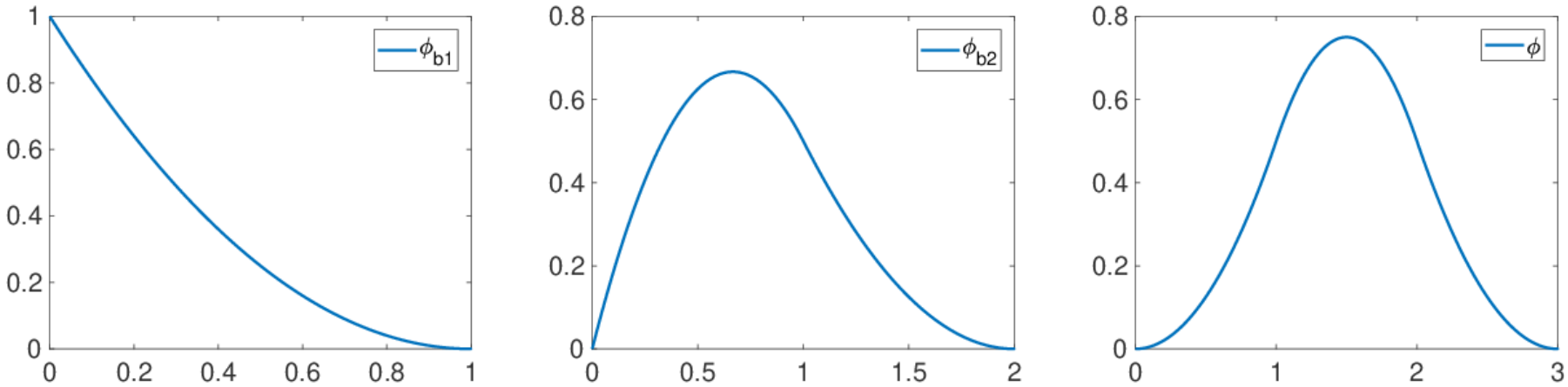

Figure: The scaling functions ϕ_{b1}, ϕ_{b2}, and ϕ.

Scaling basis

For $j \geq 2$ and $x \in [0,1]$ we set

$$\begin{aligned}
\phi_{j,k}(x) &= 2^{j/2}\phi(2^j x - k + 3), k = 3, ..., 2^j, \\
\phi_{j,1}(x) &= 2^{j/2}\phi_{b1}(2^j x), \quad \phi_{j,2^j+2}(x) = 2^{j/2}\phi_{b1}(2^j(1-x)), \\
\phi_{j,2}(x) &= 2^{j/2}\phi_{b2}(2^j x), \quad \phi_{j,2^j+1}(x) = 2^{j/2}\phi_{b2}(2^j(1-x)).
\end{aligned}$$

We denote the index sets by

$$\mathcal{I}_j = \left\{k \in \mathbb{Z} : 1 \leq k \leq 2^j + 2\right\}.$$

We define

$$\Phi_j = \{\phi_{j,k}, k \in \mathcal{I}_j\}.$$

Wavelets

We define a wavelet ψ and a boundary wavelet ψ_b as

$$\begin{aligned}\psi(x) &= -\frac{1}{4}\phi(2x) + \frac{3}{4}\phi(2x-1) - \frac{3}{4}\phi(2x-2) + \frac{1}{4}\phi(2x-3)\\ \psi_b(x) &= -\phi_{b1}(2x) + \frac{13\phi_{b2}(2x)}{12} - \frac{37\phi(2x)}{72} + \frac{\phi(2x-1)}{8}.\end{aligned}$$

Then, $\operatorname{supp}\psi = [0,3]$ and $\operatorname{supp}\psi_b = [0,2]$, i.e. the wavelets have the shortest possible support.

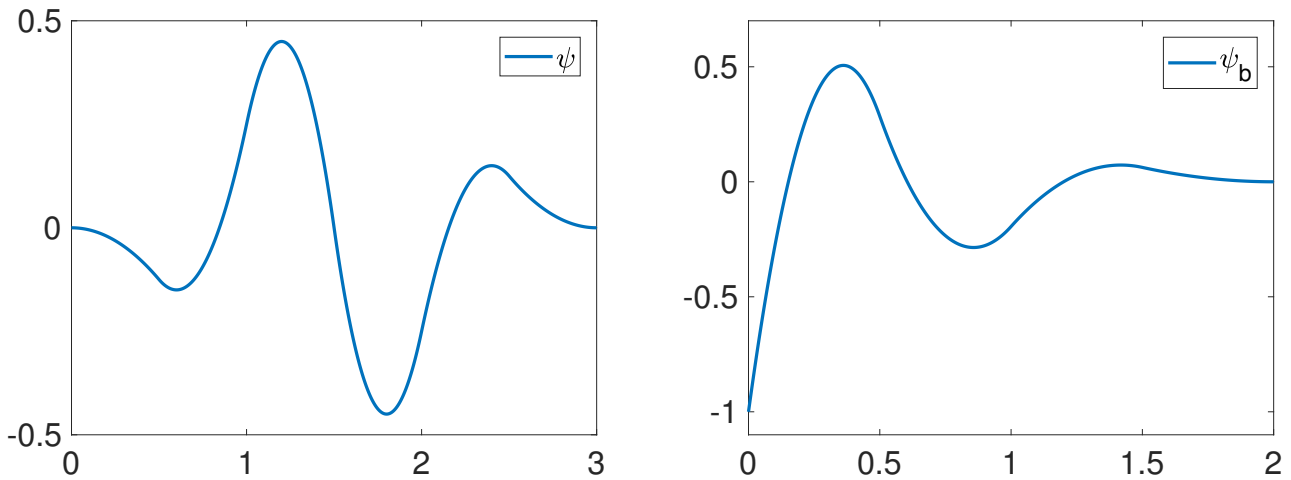

Figure: The wavelet ψ and the boundary wavelet ψ_b.

Wavelet basis

Lemma. The wavelets ψ and ψ_b have three vanishing moments, i.e.

$$\int_0^3 x^k \psi(x)dx = 0, \quad \int_0^2 x^k \psi_b(x)dx = 0, \quad k = 0, 1, 2.$$

For $j \geq 2$ and $x \in [0, 1]$ we define

$$\begin{aligned} \psi_{j,k}(x) &= 2^{j/2}\psi(2^j x - k + 2), k = 2, ..., 2^j - 1, \\ \psi_{j,1}(x) &= 2^{j/2}\psi_b(2^j x), \quad \psi_{j,2^j}(x) = 2^{j/2}\psi_b(2^j(1-x)). \end{aligned}$$

We denote the index sets by $\mathcal{J}_j = \left\{k \in \mathbb{Z} : 1 \leq k \leq 2^j\right\}$. We define

$$\Psi_j = \{\psi_{j,k}, k \in \mathcal{J}_j\},$$

and

$$\Psi = \Phi_2 \cup \bigcup_{j=2}^{\infty} \Psi_j, \quad \Psi_{j_0}^s = \Phi_{j_0} \cup \bigcup_{j=j_0}^{j_0-1+s} \Psi_j, \quad j_0 \geq 2, \quad s > 0.$$

The set Ψ_2^s is a finite-dimensional subset of Ψ. In numerical experiments we also use Ψ_3^s.

Theorem. The set Ψ is a Riesz basis of the space $L^2(0,1)$.

Corollary. The set $\Phi_{j_0} \cup \bigcup_{j=j_0}^{\infty} \Psi_j$ with the coarsest level $j_0 > 2$ is also a Riesz basis of the space $L^2(0,1)$.

Proof. The proof is long and technical. It is based on the analysis of the eigenvalues of the Gram matrices $\left\langle \Psi_{j_0}^s, \Psi_{j_0}^s \right\rangle$, because Ψ is a Riesz basis of $L^2(0,1)$ if and only if there exist constants c and C such that

$$0 < c < \lambda_{min} \left\langle \Psi_{j_0}^s, \Psi_{j_0}^s \right\rangle < \lambda_{max} \left\langle \Psi_{j_0}^s, \Psi_{j_0}^s \right\rangle < C < \infty.$$

Adaptation to homogeneous Dirichlet boundary conditions

Let ϕ, ϕ_{b2}, and ψ be defined as above and let the boundary wavelet ψ_b^D be given by

$$\psi_b^D(x) = -\frac{\phi_{b2}(2x)}{4} + \frac{47\phi(2x)}{120} - \frac{13\phi(2x-1)}{40} + \frac{\phi(2x-2)}{10}.$$

Then, ψ_b^D has three vanishing moments and $\psi_b^D(0) = 0$.

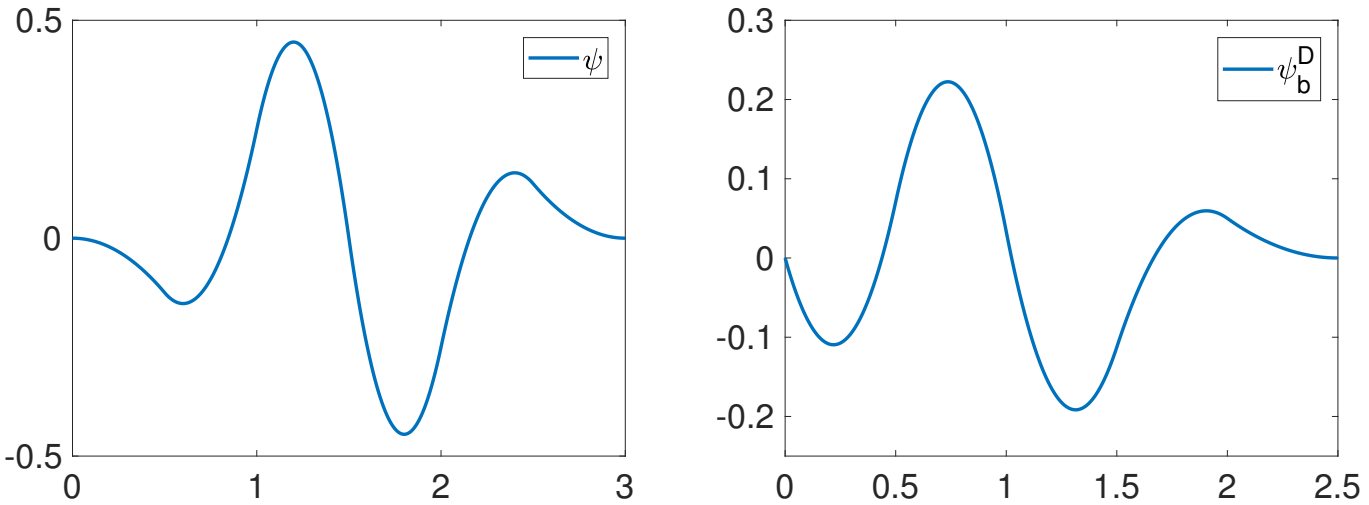

Figure: The wavelet ψ and the boundary wavelet ψ_b^D adapted to a homogeneous Dirichlet boundary condition at point $x = 0$.

Let $\phi^D_{j,k}$ be defined by

$$\begin{aligned}\phi^D_{j,k}(x) &= 2^{j/2}\phi(2^j x - k + 2), \quad k = 2, ..., 2^j - 1,\\ \phi^D_{j,1}(x) &= 2^{j/2}\phi_{b2}(2^j x), \quad \phi^D_{j,2^j}(x) = 2^{j/2}\phi_{b2}(2^j(1-x)).\end{aligned}$$

For $j \geq 2$ and $x \in [0,1]$ we define

$$\begin{aligned}\psi^D_{j,k}(x) &= 2^{j/2}\psi(2^j x - k + 2), \quad k = 2, ..., 2^j - 1,\\ \psi^D_{j,1}(x) &= 2^{j/2}\psi^D_b(2^j x), \quad \psi^D_{j,2^j}(x) = 2^{j/2}\psi^D_b(2^j(1-x)).\end{aligned}$$

We define

$$\Phi^D_j = \left\{\phi^D_{j,k}, k \in \mathcal{J}_j\right\}, \quad \Psi^D_j = \left\{\psi^D_{j,k}, k \in \mathcal{J}_j\right\},$$

and

$$\Psi^D = \Phi^D_2 \cup \bigcup_{j=2}^{\infty} \Psi^D_j, \quad \Psi^{s,D}_{j_0} = \Phi^D_{j_0} \cup \bigcup_{j=j_0}^{j_0-1+s} \Psi^D_j, \quad j_0 \geq 2.$$

Theorem. The set Ψ^D is a Riesz basis of the space $L^2(0,1)$.

Corollary. Due to this theorem, the multiscale structure of the set Ψ^D, the smoothness of functions from Ψ^D, and the polynomial exactness of Ψ^D, the set Ψ^D when normalized with respect to the H^1-norm is the Riesz basis of the space $H_0^1(0,1)$.

Construction of multidimensional wavelet basis

We constructed a wavelet basis Ψ on the interval $(0,1)$.

The wavelet basis Ψ^i on the interval $I_i = (a_i, b_i)$ is obtained by a simple linear transformation

$$\psi^i_{j,k}(x) = \psi_{j,k}\left(\frac{x - a_i}{b_i - a_i}\right), \quad x \in [a_i, b_i],$$

and similarly for the scaling functions.

We obtain the wavelet basis on the hyperrectangle $\Omega = (a_1, b_1) \times (a_2, b_2) \times \ldots \times (a_d, b_d)$, using an anisotropic tensor product, i.e. the wavelet basis on Ω is given by $\Psi = \Psi^1 \otimes \ldots \otimes \Psi^d$. We denote its subset containing s levels of wavelets starting from the coarsest level j_0 by $\Psi^s_{j_0}$.

Similarly, we construct the set Ψ^D and the finite-dimensional set $\Psi^{s,D}_{j_0}$.

Theorem. The set $\Psi = \{\psi_\lambda, \lambda \in \mathcal{J}\}$ is a Riesz basis of the space $L^2(\Omega)$ and the set Ψ^D normalized in the H^1-norm is a Riesz basis of the space $H^1_0(\Omega)$.

Wavelet-Galerkin method

Let $\Psi_{j_0}^k$ and $\Psi_{j_0}^{k,D}$ be multiscale bases. For the fixed coarsest level $j_0 \geq 2$, let us denote

$$\Psi^k = \begin{cases} \Psi_{j_0}^k, & \epsilon = 0, \\ \Psi_{j_0}^{k,D}, & \epsilon > 0. \end{cases}$$

Thus Ψ^k is a wavelet basis that contains scaling functions at a coarsest level j_0 and k levels of wavelets. Then $X_k = \operatorname{span} \Psi^k$ are the finite-dimensional subspaces of V that are nested, i.e. $X_k \subset X_{k+1}$, $k \in \mathbb{N}$, and

$$V = \overline{\bigcup_{k=j_0-1}^{\infty} X_k}.$$

The Galerkin formulation of (1) reads: Find $y_k \in X_k$ such that

$$a(y_k, v) = \langle f, v \rangle \quad \text{for all} \quad v \in X_k. \tag{3}$$

Theorem. Let (A1)–(A4) be satisfied, $h = 2^{-j_0-k}$ and y be a solution of (1). Then, there exists a unique solution y_k of Equation (3) and $\|y_k - y\|_V \to 0$. Moreover, if $y \in H^3(\Omega) \cap V$, then

$$\|y_k - y\| \leq Ch^3, \quad \|y_k - y\|_{H^1} \leq Ch^2$$

with a constant C that does not depend on h.

Proof. The existence and uniqueness of the solution y_k of the problem (3) is a consequence of the Lax-Milgram lemma. Let $\mathcal{A}$ be defined as above, $\mathcal{P}_k : V \to X_k$ be a V-orthogonal projection, and

$$\mathcal{A}_k := \mathcal{P}_k \mathcal{A}|_{X_k}.$$

It is known that under the assumptions that $\mathcal{A}$ is invertible, there exists $C \in \mathbb{R}$ such that

$$\left\| \mathcal{A}_k^{-1} \mathcal{P}_k \mathcal{A} \right\| \leq C,$$

and that the denseness property is satisfied, it holds that $\|y_k - y\|_V \to 0$ and

$$\|y_k - y\|_V \leq C \|\mathcal{P}_k y - y\|_V .$$

We verify these assumptions. As usual, C denotes a generic constant that may take different values at different occurrences. The invertibility of $\mathcal{A}$ and $\mathcal{A}_k$ and boundedness of the norm of $\mathcal{A}_k^{-1}$ is a consequence of the Lax-Milgram lemma.

Moreover, due to the orthogonality of the projections $\mathcal{P}_k$ and boundedness of $\mathcal{A}$ we have

$$\left\|\mathcal{A}_k^{-1}\right\| \leq \frac{1}{\alpha}, \quad \|\mathcal{P}_k\| = 1, \quad \|\mathcal{A}\| \leq \beta,$$

where α is a coercivity constant.
Hence, the error depends on the approximation properties of spline spaces X_k, which are well-known. For $y \in H^3(\Omega) \cap V$, we have

$$\|\mathcal{P}_k y - y\|_{H^1} \leq Ch^{-1} \|\mathcal{P}_k y - y\| \leq Ch^2 \|y\|_{H^3}.$$

Stability of the solution

Theorem. Let $\mathcal{P}_k : V \to X_k$ be a V-orthogonal projection and $\mathcal{A}_k$ be defined by

$$\mathcal{A}_k := \mathcal{P}_k \mathcal{A}|_{X_k}.$$

The above method is stable in the sense that there exist nonnegative constants μ and ν, a positive constant δ and a positive integer q such that for any $k \geq q$, the operator $\mathcal{A}_k$ is invertible, and the perturbed equation $(\mathcal{A}_k + \mathcal{E}_k)\,\tilde{y}_k = \mathcal{P}_k f + g_k$ has a unique solution $\tilde{y}_k \in X_k$ for any $g_k \in X_k$ and any bounded linear operator $\mathcal{E}_k : X_k \to X_k$, with $\|\mathcal{E}_k\| < \delta$, and $\tilde{y}_k$ satisfies

$$\|\tilde{y}_k - y_k\|_V \leq \mu \|\mathcal{E}_k\| \|y_k\|_V + \nu \|g_k\|_V .$$

From the above Theorem, the convergence rate depends on the chosen discretization spaces and not directly on the chosen bases of these spaces.

Since the constructed basis generates the same spaces as quadratic B-splines or other quadratic spline wavelets, it can be expected that the error will be similar. The main difference is therefore in the sparsity of the discretization matrices, the condition numbers of the matrices and the number of iterations needed to resolve the problem with a desired accuracy.

We write the function y_k as

$$y_k = \sum_{\psi_\lambda \in \Psi^k} c_\lambda^k \psi_\lambda .$$

Let $\mathbf{G}^k$ and $\mathbf{K}^k$ be matrices with the entries

$$\mathbf{G}^k_{\mu,\lambda} = \epsilon \left\langle \psi_\lambda, \psi_\mu \right\rangle + \left\langle p\psi_\lambda, \psi_\mu \right\rangle , \quad \mathbf{K}^k_{\mu,\lambda} = \left\langle \mathcal{K}\psi_\lambda, \psi_\mu \right\rangle ,$$

for $\psi_\lambda, \psi_\mu \in \Psi^k$. Let $\mathbf{f}^k$ be a vector with entries

$$f^k_\mu = \left\langle f, \psi_\mu \right\rangle , \quad \psi_\mu \in \Psi^k,$$

and $\mathbf{c}^k$ be the column vector of coefficients c^k_λ. We obtain the system

$$\mathbf{A}^k \mathbf{c}^k = \mathbf{f}^k, \quad \text{where} \quad \mathbf{A}^k = \mathbf{G}^k + \mathbf{K}^k.$$

We apply the standard Jacobi diagonal preconditioning. Let $\mathbf{D}^k$ be a diagonal matrix with diagonal elements

$$\mathbf{D}^k_{\lambda,\lambda} := \sqrt{\mathbf{A}^k_{\lambda,\lambda}} = \sqrt{a(\psi_\lambda, \psi_\lambda)}.$$

We define the preconditioned system

$$\tilde{\mathbf{A}}^k \tilde{\mathbf{c}}^k = \tilde{\mathbf{f}}^k$$

with

$$\tilde{\mathbf{A}}^k = \left(\mathbf{D}^k\right)^{-1} \mathbf{A}^k \left(\mathbf{D}^k\right)^{-1}, \quad \tilde{\mathbf{f}}^k = \left(\mathbf{D}^k\right)^{-1} \mathbf{f}^k, \quad \tilde{\mathbf{c}}^k = \mathbf{D}^k \mathbf{c}^k.$$

We solve this system by the method of generalized residuals (GMRES) or, in the case where the system matrix is symmetric and positive definite, we use the conjugate gradient method.

Uniform boundedness of the condition numbers

Theorem. There exists a constant C such that

$$\operatorname{cond} \tilde{\mathbf{A}}^k \leq C$$

for all $k \geq 0$.

Proof. The proof is quite long and technical, it is based on the continuity and coercivity of the bilinear form a and the Riesz basis property of Ψ.

Compression of the discretization matrices

We study the structure of the matrices $\mathbf{G}^k$ and $\mathbf{K}^k$. Let the size of these matrices be $N \times N$ (with N dependent on k). The matrix $\mathbf{G}^k$ has $\mathcal{O}(N \ln N)$ nonzero entries and has a so-called finger-band pattern. For $\epsilon = 0$, $p = 1$, and the constructed wavelet basis, the pattern is displayed below. For seven levels of wavelets the matrix has the size 514×514 and thus it has 264196 entries, but only 16294 of the entries are nonzero.

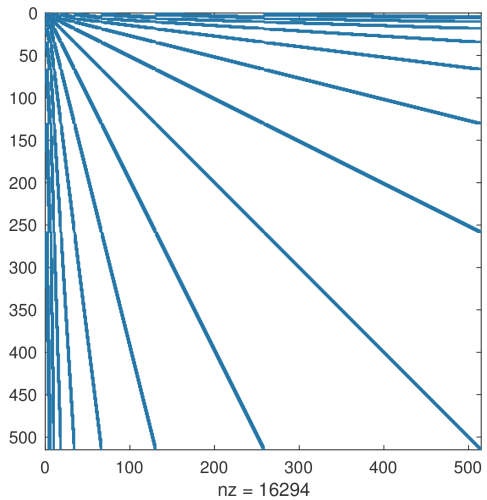

Figure: The sparsity pattern of the matrix $\mathbf{G}^k$ for seven levels of wavelets.

Decay estimates

For the standard Galerkin method using B-splines, the matrix $\mathbf{K}^k$ is full. However, it is known that for some classes of operators and some types of bases, many entries of the matrix $\mathbf{K}^k$ are small and can be thresholded. Consequently, the matrix $\mathbf{K}^k$ can be approximated with a matrix that is sparse. [Beylkin et al. 1992]

The decay estimates of the entries of the matrix $\mathbf{K}^k$ are typically derived for isotropic systems. We present the decay estimates for anisotropic wavelet systems and a kernel K that is sufficiently smooth.

Theorem. Let $\psi_\lambda, \psi_\mu \in \Psi^k$ be wavelets with $L = 3$ vanishing moments and let the assumption (A2) be fulfilled for $m = 2L$. Then,

$$\left| \int_\Omega \int_\Omega K(x,t)\, \psi_\lambda(x)\, \psi_\mu(t)\, dx\, dt \right| \leq C 2^{-([\lambda]+[\mu])(L+d/2)},$$

with a constant C independent of λ and μ and $[\lambda]$ is the level of ψ_λ.

Proof. Let the centres of the supports of ψ_λ and ψ_μ be denoted by x_λ and t_λ, respectively. For multi-index $l = (l_1, \ldots, l_d) \in \mathbb{N}_0^d$, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, and $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, we denote $|l| = l_1 + \ldots + l_d$, $l! = l_1! \ldots l_d!$, and $x^l = x_1^{l_1} \ldots x_d^{l_d}$. Due to (A2) and the Taylor Theorem, there exists a function P such that for fixed t the function $P(x,t)$ is a polynomial with respect to x of degree at most $L-1$, and there exists a function Q such that for fixed x the function $Q(x,t)$ is a polynomial with respect to t of degree at most $L-1$ such that

$$K(x,t) = P(x,t) + Q(x,t) + \sum_{|l|=L} \sum_{|m|=L} C_{l,m} (x - x_\lambda)^l (t - t_\mu)^m \quad (4)$$

with

$$C_{l,m} = \frac{1}{l!m!} \frac{\partial^l \partial^m K(\xi(x,t))}{\partial x^l \partial t^m} \quad (5)$$

and

$$\xi(x,t) = (x_\lambda, t_\mu) + \alpha((x,t) - (x_\lambda, t_\mu)) \quad (6)$$

for some $\alpha \in [0,1]$.

Since Ψ has L vanishing moments, we have

$$\int_{\Omega}\int_{\Omega}\left(P\left(x,t\right)+Q\left(x,t\right)\right)\psi_{\lambda}\left(x\right)\psi_{\mu}\left(t\right)dx\,dt=0. \tag{7}$$

Let us denote

$$K_{\mu,\lambda}=\int_{\Omega}\int_{\Omega}K\left(x,t\right)\psi_{\lambda}\left(x\right)\psi_{\mu}\left(t\right)dx\,dt. \tag{8}$$

If $|l|=L$ then

$$\left|x-x_{\lambda}\right|^{l}\leq C\prod_{i=1}^{d}2^{-|\lambda_i|l_i}\leq C2^{-L[\lambda]} \tag{9}$$

and

$$\int_{\Omega}\left|\psi_{\lambda}\left(x\right)\right|dx\leq C2^{-|\lambda_1|/2-\ldots-|\lambda_d|/2}\leq C2^{-[\lambda]d/2}. \tag{10}$$

Hence combining (4), (7), (9), and (10), we have

$$\begin{aligned} |K_{\mu,\lambda}| &\leq \sum_{|l|=L} \sum_{|m|=L} C_{l,m} \int_{\Omega} \int_{\Omega} |x - x_\lambda|^l \, |t - t_\mu|^m \, |\psi_\lambda(x) \, \psi_\mu(t)| \, dx \, dt \\ &\leq C 2^{-L[\lambda]-L[\mu]-[\lambda]d/2-[\mu]d/2}. \end{aligned}$$

Let $\tilde{\mathbf{A}}^k$ be the matrix defined above. Due to derived decay estimates and the local support of wavelets, many entries of the matrix $\tilde{\mathbf{A}}^k$ are small and they can be thresholded, and the matrix $\tilde{\mathbf{A}}^k$ can be approximated by sparse matrix $\hat{\mathbf{A}}^k$.

More precisely, let T be a chosen threshold and let $\hat{\mathbf{A}}^k$ be defined as

$$\hat{\mathbf{A}}^k_{m,l} = \begin{cases} \tilde{\mathbf{A}}^k_{m,l}, & \text{if } \left|\tilde{\mathbf{A}}^k_{m,l}\right| > T, \\ 0, & \text{if } \left|\tilde{\mathbf{A}}^k_{m,l}\right| \leq T. \end{cases}$$

Then,

$$\left\|\tilde{\mathbf{A}}^k - \hat{\mathbf{A}}^k\right\| \leq \max_m \sum_l \left|\left(\tilde{\mathbf{A}}^k - \hat{\mathbf{A}}^k\right)_{m,l}\right| \leq T n_k,$$

where $n_k \times n_k$ is the size of the matrix $\tilde{\mathbf{A}}^k$.

Theorem. If $\tilde{\mathbf{c}}^k$ is the solution of the system with the matrix $\tilde{\mathbf{A}}^k$, and $\hat{\mathbf{c}}^k$ is the solution of the system with the matrix $\hat{\mathbf{A}}^k$ and the same right-hand side, and if

$$\gamma := \left\| \left(\tilde{\mathbf{A}}^k \right)^{-1} \left(\tilde{\mathbf{A}}^k - \hat{\mathbf{A}}^k \right) \right\| < 1,$$

then

$$\frac{\left\| \tilde{\mathbf{c}}^k - \hat{\mathbf{c}}^k \right\|}{\left\| \hat{\mathbf{c}}^k \right\|} \leq \frac{\operatorname{cond} \tilde{\mathbf{A}}^k}{1-\gamma} \frac{\left\| \tilde{\mathbf{A}}^k - \hat{\mathbf{A}}^k \right\|}{\left\| \tilde{\mathbf{A}}^k \right\|} \leq \frac{T n_k \operatorname{cond} \tilde{\mathbf{A}}^k}{(1-\gamma) \left\| \tilde{\mathbf{A}}^k \right\|}.$$

Moreover, the condition numbers and the norms of the matrices $\tilde{\mathbf{A}}^k$ are uniformly bounded.

Compression strategy.

Strategy 1.
In some applications where the left-hand side of the equation is fixed and the equation is solved for various right-hand sides, the system matrix can be computed, analysed and compressed only once as a preprocessing step and then one can work with the compressed matrix.

Strategy 2. Another approach is to use the derived decay estimate together with the known structure of the matrices $\mathbf{G}^k$ to compute only significant entries of the matrix $\mathbf{A}^k$.

Algorithm.

1. Compute the significant entries of the matrix $\tilde{\mathbf{A}}^k$ and the vector of the right-hand side $\tilde{\mathbf{f}}^k$.
2. Solve the system $\tilde{\mathbf{A}}^k \tilde{\mathbf{c}}^k = \tilde{\mathbf{f}}^k$.
3. Compute the solution

$$y_k = \sum_{\psi_\lambda \in \Psi^k} c_\lambda^k \psi_\lambda,$$

where c_λ^k are elements of $\mathbf{c}^k = \left(\mathbf{D}^k\right)^{-1} \tilde{\mathbf{c}}^k$.

Numerical examples - notation

e_2 denotes the L^2-norm of the error, i.e. $\|y - y_k\|$

$N \times N$ is the size of the system matrix

K denotes the finest level of a wavelet basis

NNZ is the number of nonzero entries of the system matrix

$cond$ is the condition number of the system matrix

it is the number of iterations for the chosen iterative method

$CR = NNZ/N^2$ is the compression ratio

Methods and bases

B-spline denotes the standard quadratic B-spline basis.

Quadratic semiorthogonal wavelet basis constructed by Chui and Quak is denoted as CQ.

Biorthogonal quadratic spline wavelet bases with k vanishing moments are denoted as *bior3.k*.

The constructed wavelet bases are denoted as *new*.

Spline-Galerkin denotes the Galerkin method with quadratic B-splines.

Spline-collocation denotes the collocation method with quadratic B-splines (with 2^j uniform collocation nodes for level j).

Legendre-collocation denotes the collocation method with Legendre polynomials and Gauss-Legendre nodes.

Quadrature method denotes the standard quadrature method using the Simpson's rule.

Example 1. One-dimesional integral equation

We consider the integral equation

$$(2-t)\,y(t)+\frac{1}{\pi}\int_0^1 \sin(t-x)\,y(x)\,dx=f(t),\quad t\in(0,1),$$

with the oscillatory solution $y(t)=\sin(120\pi t)$.

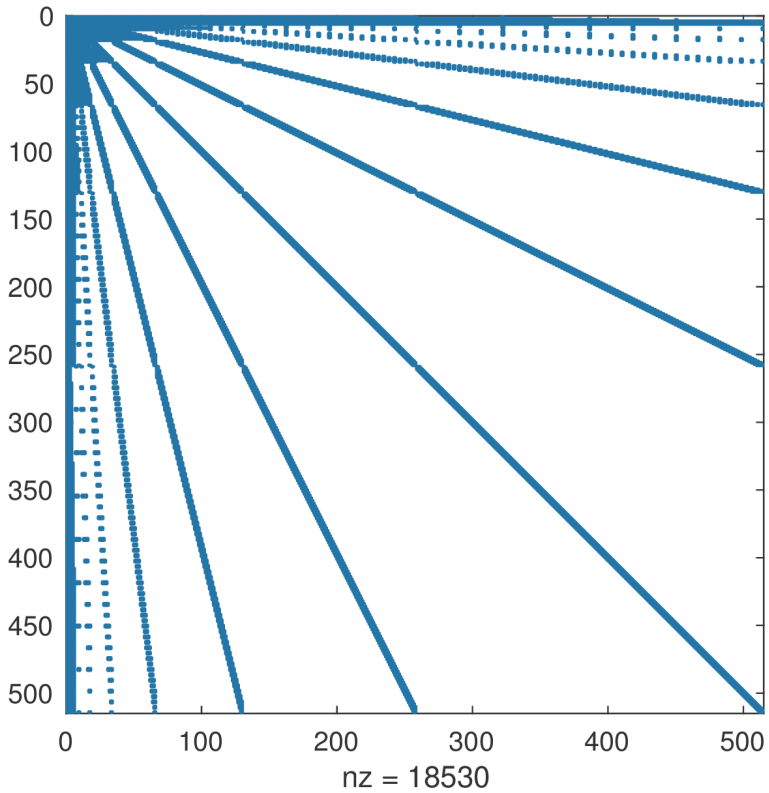

Figure: The sparsity pattern of the discretization matrix for Example 1.

		$T = 10^{-8}$		$T = 10^{-10}$		uncompressed	
K	N	e_2	NNZ	e_2	NNZ	e_2	NNZ
6	66	7.04e-1	1780	7.04e-1	1814	7.04e-1	4356
7	130	4.01e-1	3992	4.01e-1	4398	4.01e-1	16900
8	258	2.06e-2	9204	2.06e-2	10274	2.06e-2	66564
9	514	8.71e-4	21182	8.71e-4	22262	8.71e-4	264196
10	1026	2.02e-4	44052	2.02e-4	49346	2.02e-4	1052676
11	2050	2.37e-5	106396	2.37e-5	111744	2.37e-5	4202500

Table: L^2 errors and the number of nonzero entries of compressed and uncompressed matrices for Example 1.

The L^2 errors for the Galerkin method with B-splines were (to at least three decimal digits) same as the errors for the new method, as was the case for the Galerkin method with semiorthogonal wavelets and bior3.3 wavelets.

The results are also displayed for the spline collocation method, the Legendre collocation method, and the quadrature method. All four methods lead to matrices that are full and in the case of Legendre collocation method they are also very badly conditioned.

Spline-Galerkin			*Spline-collocation*		
N	e_2	cond	N	e_2	cond
66	7.04e-1	7.64e0	66	1.00e0	4.12e0
130	4.01e-1	7.62e0	130	4.05e-1	4.17e0
258	2.06e-2	7.62e0	258	2.51e-2	4.20e0
514	8.71e-4	7.62e0	514	2.01e-3	4.22e0
1026	2.02e-4	7.61e0	1026	2.15e-4	4.23e0
2050	2.37e-5	7.61e0	2050	2.48e-5	4.23e0
Legendre-collocation			*Quadrature method*		
N	e_2	cond	N	e_2	cond
66	1.80e1	1.74e18	67	9.82e-1	2.00e0
130	1.81e1	3.99e18	151	3.54e-1	2.00e0
258	1.51e1	2.69e18	259	1.32e-1	2.00e0
514	1.21e-13	5.18e19	515	6.63e-2	2.00e0
1026	1.61e-12	1.14e20	1027	8.69e-3	2.00e0

	B-spline		*CQ*		*bior*3.3		*new*	
K	*NNZ*	*cond*	*NNZ*	*cond*	*NNZ*	*cond*	*NNZ*	*cond*
3	100	7.9	100	8.4	100	7.9	100	8.8
4	324	7.7	312	7.9	316	25.1	286	8.9
5	1156	7.6	932	7.9	956	37.4	726	8.9
6	4356	7.6	2552	7.9	2604	53.0	1814	8.9
7	16900	7.6	6494	7.9	6628	64.3	4398	8.9
8	66564	7.6	14002	7.9	16012	75.0	10274	8.9
9	264196	7.6	31524	7.9	34060	84.0	22262	8.9
10	1052676	7.6	70410	7.9	75380	92.0	49346	8.9
11	4202500	7.6	162190	7.9	168386	98.9	111744	8.9

Table: The number of nonzero entries and the condition numbers of compressed matrices with threshold 10^{-10} for various bases for Example 1.

In the table, the number of nonzero entries and the condition numbers of the compressed matrices with the threshold $T = 10^{-10}$ are listed for various piecewise quadratic bases. We also tested biorthogonal quadratic spline wavelet bases with 5 and 7 vanishing moments, but the results were significantly worse than for bases listed in the table both with respect to the number of nonzero entries and the condition number of the compressed discretization matrix.

The number of nonzero entries of the system matrix is the smallest for the new basis.

Example 2. Two-dimensional integral equation

We consider $\Omega = (0,1)^2$ and the equation

$$p(t_1,t_2)\,y(t_1,t_2) + \iint_\Omega K(x_1,x_2)\,y(x_1,x_2)\,dx_1\,dx_2 = f(t_1,t_2),$$

where

$$p(t_1,t_2) = (t_1+0.1)^2\,(1.1+\sin 10\pi t_2),\; K(x_1,x_2) = (x_1 \sin x_2 + 1),$$

$(t_1,t_2) \in \Omega$, and the solution is $y(t_1,t_2) = t_1 \cos(50\pi t_2)$. We present the results for the basis Ψ_3^{K-3}, because they were slightly better than then results for the basis Ψ_2^{K-2}.

We solve the resulting system by the GMRES method. The results for bases *CQ* and *new* are presented. Since the Galerkin method with B-splines, the spline-collocation method, the Legendre-collocation method and the quadrature method lead to full matrices, we do not use them for this problem. The Galerkin method with bior3.3 wavelets leads to badly conditioned matrices and the GMRES method stopped after 500 iterations without reaching the desired residual.

	CQ			*new*		
N	e_2	it	CR	e_2	it	CR
1156	4.10e-1	43(9)	2.67e-1	4.10e-1	46(3)	1.17e-1
4356	1.06e-1	42(7)	1.31e-1	1.06e-1	35(7)	8.23e-2
16900	6.12e-3	43(2)	4.82e-2	6.12e-3	35(1)	3.15e-2
66564	5.95e-4	43(2)	1.36e-2	5.95e-4	34(1)	9.43e-3
264196	6.70e-5	43(2)	3.25e-3	6.70e-5	34(10)	2.44e-3

Table: The compression ratios for $T = 10^{-8}$, the number of outer and inner GMRES iterations and the L^2 errors for Example 2.

Example 3. One-dimensional integro-differential equation I.

We use the Galerkin method with the basis $\Psi_2^{K-2,D}$ for the numerical solution of the integro-differential equation

$$-y''(t)+y(t)-\frac{1}{2}\int_0^1 \sin(\pi x+\pi t)\,y(x)\,dx=f(t),\quad t\in(0,1),$$

with the boundary conditions $y(0)=y(1)=0$. The right-hand side f is such that the solution is $y(t)=\sin(40\pi t)$. We use the threshold $T=10^{-10}$ for matrix compression. Since the CQ basis can not be adapted to boundary conditions while preserving both semiorthogonality and the number of vanishing moments, and boundary adapted bior3.5 wavelets led to better results than boundary adapted bior3.3 wavelets, we present in Table 5 the results for the Galerkin method with bases *bior*3.5 and *new*.

	bior3.5			*new*		
N	e_2	cond	CR	e_2	cond	CR
64	7.19e-2	13.03	6.34e-1	7.19e-2	10.77	3.94e-1
128	4.97e-3	13.16	4.00e-1	4.97e-3	10.86	1.98e-1
256	5.15e-4	13.20	2.32e-1	5.15e-4	10.90	9.96e-2
512	6.12e-5	13.23	1.28e-1	6.12e-5	10.93	5.05e-2
1024	7.53e-6	13.23	6.77e-2	7.53e-6	10.95	2.56e-2
2048	9.03e-7	13.23	3.51e-2	9.03e-7	10.96	1.29e-2

Table: The compression ratios, L^2 errors and the condition numbers for Example 3.

For these methods the matrices were full and badly conditioned.

	Spline-Galerkin		*Spline-collocation*	
N	e_2	cond	e_2	cond
64	7.19e-2	5.65e2	1.22e-1	1.51e3
128	4.97e-3	2.26e3	2.84e-2	6.03e3
256	5.15e-4	9.05e3	7.10e-3	2.41e4
512	6.12e-5	3.62e4	1.77e-3	9.65e4
1024	7.72e-6	1.44e5	4.44e-4	3.86e5

Table: The condition numbers and L^2 errors for Example 3.

Example 4. One-dimensional integro-differential equation II.

We solve the integro-differential equation

$$-y''(t) + 2y(t) - \int_0^1 (x-t)\, y(x)\, dx = f(t), \quad t \in (0,1),$$

with the boundary conditions $y(0) = y(1) = 0$. The right-hand side f is such that the solution is $y(t) = (1-t)\left(1-e^{10t}\right)$. We use the basis $\Psi_2^{K-2,D}$ and the threshold $T = 10^{-10}$. We use the same methods as in the previous example.

	bior3.5			*new*		
N	e_2	cond	CR	e_2	cond	CR
64	2.78e-2	12.1	6.08e-1	2.78e-2	10.8	3.10e-1
128	3.45e-3	12.1	3.94e-1	3.45e-3	10.8	1.74e-1
256	4.31e-4	12.2	2.31e-1	4.31e-4	10.9	9.35e-2
512	5.38e-5	12.2	1.27e-1	5.38e-5	10.9	4.90e-2
1024	6.71e-6	12.2	6.76e-2	6.71e-6	10.9	2.52e-2
2048	8.08e-7	12.2	3.51e-2	8.08e-7	10.9	1.28e-2

Table: The compression ratios, L^2 errors and the condition numbers for Example 4.

	Spline-Galerkin		*Spline-collocation*	
N	e_2	cond	e_2	cond
64	2.78e-2	5.18e2	2.00e0	1.38e3
128	3.45e-3	2.07e3	5.01e-1	5.52e3
256	4.31e-4	8.28e3	1.25e-1	2.21e4
512	5.38e-5	3.31e4	3.14e-2	8.33e4
1024	6.71e-6	1.33e5	7.84e-3	3.53e5

Table: The condition numbers and L^2 errors for Example 4.

Example 5. Two-dimesional integro-differential equation

For $\Omega = (0,1)^2$ we consider the equation

$$-\epsilon \Delta y(t_1,t_2) + y(t_1,t_2) + \iint_\Omega \frac{e^{x_1+t_1} x_2 t_2}{2} y(x_1,x_2) = f(t_1,t_2),$$

with the homogeneous Dirichlet boundary conditions, $\epsilon = 10^{-5}$, and with the solution $y(t_1,t_2) = t_1 t_2 \left(1 - e^{50t_1-50}\right)\left(1 - e^{50t_2-50}\right)$.

We use $\Psi_3^{K-3,D}$ basis and the threshold $T = 10^{-10}$. For this equation the system matrix is symmetric and positive definite and thus we use the conjugate gradient method. The iterations stop if the relative residual is less than 10^{-10}.

	bior3.5			*new*		
N	e_2	it	CR	e_2	it	CR
1024	1.97e-3	300	6.23e-1	1.97e-3	104	2.06e-1
4096	2.41e-4	419	3.60e-1	2.41e-4	134	8.16e-2
16384	2.88e-5	488	1.45e-1	2.88e-5	157	3.00e-2
65536	3.54e-6	514	4.85e-2	3.54e-6	171	8.24e-3
262144	4.89e-7	520	1.43e-2	4.89e-7	176	2.31e-3

Table: The compression ratios for $T = 10^{-9}$, the number conjugate gradient iterations and L^2 errors for Example 5.

Conclusions

- We constructed quadratic spline-wavelet bases with short supports and with three vanishing moments for the spaces $L^2(\Omega)$ and $H_0^1(\Omega)$, where Ω is the hyperrectangle.
- We used the Galerkin method with the constructed bases for solving integral and integro-differential equations. The method is convergent and stable and the discretization matrices have uniformly bounded condition numbers.
- Based on the decay estimates of the elements of discretization matrices, the compression strategy was proposed.
- We presented several numerical examples and compared the results with the Galerkin method with other quadratic spline wavelet bases and with other methods.
- The errors for small enough thresholds were similar as the error for full matrices, but the number of nonzero entries of the compressed matrices was significantly smaller than for full matrices and thus we were able to solve large systems efficiently.

References

- Beylkin, G., Coifman, R., Rokhlin, V.: Fast wavelet transforms and numerical algorithms I. Communications on Pure and Applied Mathematics 44, 141Â–183 (1991).
- Černá, D., Finěk, V.: Construction of optimally conditioned cubic spline wavelets on the interval. Adv. Comput. Math. 34, 219–252 (2011).
- D. Černá, V. Finěk: Galerkin method with new quadratic spline wavelets for integral and integro-differential equations. *Journal of Computational and Applied Mathematics* **363**, 2020, pp. 426-443.
- Chui, C.K., Quak, E.: Wavelets on a bounded interval. In: Braess, D., Schumaker, L.L. (eds.), Numerical Methods of Approximation Theory, pp. 53–75, Birkhäuser (1992).
- Dahmen, W., Kunoth, A.: Multilevel Preconditioning. Numer. Math. 63, 315–344 (1992).