Journal of Computational and Applied Mathematics 260 (2014) 103–116

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Fast evaluation of system matrices w.r.t. multi-tree collections of tensor product refinable basis functions

Sebastian Kestler [a,*], Rob Stevenson [b]

[a] Institute for Numerical Mathematics, University of Ulm, Helmholtzstrasse 20, D-89069 Ulm, Germany

[b] Korteweg–de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 8 March 2012
Received in revised form 8 November 2012

MSC:
41A30
65N30
65Y20

Keywords:
Tensor product approximation
Adaptivity
Sparse grids
Optimal computational complexity

ABSTRACT

An algorithm is presented that for a local bilinear form evaluates in linear complexity the application of the stiffness matrix w.r.t. a collection of tensor product multiscale basis functions, assuming that this collection has a multi-tree structure. It generalizes an algorithm for sparse-grid index sets [R. Balder, Ch. Zenger, The solution of multidimensional real Helmholtz equations on sparse grids, SIAM J. Sci. Comput. 17 (3) (1996) 631–646] and it finds its application in adaptive tensor product approximation methods.

1. Introduction

For $1 \le i \le n$, let $\check{\Psi}_i = \{\check{\psi}_{i,\lambda} : \lambda \in \check{\nabla}_i\}$ and $\hat{\Psi}_i = \{\hat{\psi}_{i,\lambda} : \lambda \in \hat{\nabla}_i\}$ be collections of multi-scale functions, e.g., wavelet bases, multi-level frames, or collections of hierarchical "hat" functions. We assume that the functions from these collections, which we simply will refer to as being wavelets, satisfy standard locality assumptions meaning that the diameter of the support of wavelets on *level* ℓ is of order $2^{-\ell}$. Let $a_i(\cdot,\cdot)$ be a bilinear form that is *local* meaning that $a_i(u,v) = 0$ whenever $|\operatorname{supp} u \cap \operatorname{supp} v| = 0$.

We set $\check{\mathbf{\Psi}} = \{\check{\boldsymbol{\psi}}_{\boldsymbol{\lambda}} = \otimes_{i=1}^n \check{\psi}_{\lambda_i} : \boldsymbol{\lambda} \in \check{\boldsymbol{\nabla}} := \prod_{i=1}^n \check{\nabla}_i\}$ and similarly $\hat{\mathbf{\Psi}}$, and consider the bilinear form $\mathbf{a}(\cdot,\cdot)$ defined by $\mathbf{a}(\otimes_{i=1}^n u_i, \otimes_{i=1}^n v_i) = \prod_{i=1}^n a_i(u_i, v_i)$. A typical example being $\mathbf{a} = \mathbf{a}_k$ defined by $a_i(u_i,v_i) = \begin{cases} \int_0^1 u_i(x)v_i(x)dx, & i \neq k, \\ \int_0^1 u_i'(x)v_i'(x)dx, & i = k, \end{cases}$ in which case $\sum_{k=1}^n \mathbf{a}_k$ is the bilinear form that results from the variational formulation of Poisson's problem on $(0,1)^n$.

The topic of this paper is to apply, for finite $\check{\mathbf{\Lambda}} \subset \check{\boldsymbol{\nabla}}$, $\hat{\mathbf{\Lambda}} \subset \hat{\boldsymbol{\nabla}}$, the "system matrix"

$$\mathbf{a}(\check{\mathbf{\Psi}}|_{\check{\mathbf{\Lambda}}}, \hat{\mathbf{\Psi}}|_{\hat{\mathbf{\Lambda}}}) := [\mathbf{a}(\check{\boldsymbol{\psi}}_{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}_{\boldsymbol{\mu}})]_{\boldsymbol{\lambda}\in\check{\mathbf{\Lambda}},\boldsymbol{\mu}\in\hat{\mathbf{\Lambda}}} \quad \text{in } \mathcal{O}(\#\check{\mathbf{\Lambda}} + \#\hat{\mathbf{\Lambda}}) \text{ operations.} \tag{1.1}$$

For doing so, there is no real restriction to assume that the collections $\check{\Psi}_i$ and $\hat{\Psi}_i$, and the bilinear forms $a_i(\cdot,\cdot)$ are independent of i. Furthermore, inside this introduction we focus on the simplified case where $(\Psi =)\check{\Psi} = \hat{\mathbf{\Psi}}$ and $(\mathbf{\Lambda} =)\check{\mathbf{\Lambda}} = \hat{\mathbf{\Lambda}}$.

* Corresponding author. Tel.: +49 731 50 23935; fax: +49 731 50 23548.
E-mail addresses: sebastian.kestler@gmail.com, sebastian.kestler@uni-ulm.de (S. Kestler), r.p.stevenson@uva.nl (R. Stevenson).

0377-0427/$ – see front matter
http://dx.doi.org/10.1016/j.cam.2013.09.015

Already for $n = 1$, the application of the system matrix in linear complexity cannot be expected for arbitrary $\Lambda \subset \nabla$, because $a(\Psi|_\Lambda, \Psi|_\Lambda)$ is generally not sparse due to interactions between wavelets on different levels. Yet, when Λ is the collection of indices of *all* wavelets up to some level ℓ, a solution is provided by the application of a transformation T to a single-scale basis Φ_ℓ on level ℓ. Writing $\Psi|_\Lambda^\top = \Phi_\ell^\top T$, viewing collections of functions as column vectors, we have $a(\Psi|_\Lambda, \Psi|_\Lambda) = T^\top a(\Phi_\ell, \Phi_\ell)T$. By the sparsity of $a(\Phi_\ell, \Phi_\ell)$, and under the assumption that each of its entries can be computed in $\mathcal{O}(1)$ operations, each of these three matrices on the right-hand side can be applied in $\mathcal{O}(\#\Lambda)$ operations, and so can $a(\Psi|_\Lambda, \Psi|_\Lambda)$.

This approach extends to the situation where Λ is a general *tree*, which we define as a set such that for any $\lambda \in \Lambda$ with $|\lambda| > 0$, the support of ψ_λ is covered by the supports of ψ_μ for some $\mu \in \Lambda$ with $|\mu| = |\lambda| - 1$. The argument is that for a tree Λ, there exists a locally finite collection of scaling functions whose span contains span $\Psi|_\Lambda$, where, thanks to the tree constraint, the representation of the embedding, being a generalization the aforementioned basis transformation T, can be performed in $\mathcal{O}(\#\Lambda)$ operations.

For $n > 1$, we have

$$\mathbf{a}(\boldsymbol{\Psi}|_{\boldsymbol{\Lambda}}, \boldsymbol{\Psi}|_{\boldsymbol{\Lambda}}) = R_{\boldsymbol{\Lambda}}(A \otimes \cdots \otimes A)I_{\boldsymbol{\Lambda}},$$

where $A := a(\Psi, \Psi)$, $I_{\boldsymbol{\Lambda}}$ is the extension operator with zeros of a vector indexed by $\boldsymbol{\Lambda}$ to one indexed by $\boldsymbol{\nabla}$, and $R_{\boldsymbol{\Lambda}}$ denotes its adjoint being the restriction of a vector to its indices in $\boldsymbol{\Lambda}$.

If $\boldsymbol{\Lambda}$ is equal to $\bar{\boldsymbol{\Lambda}} := \{\boldsymbol{\lambda} \in \boldsymbol{\nabla} : \||\boldsymbol{\lambda}|\|_\infty \le \ell\}$, being the set of all multi-indices $\boldsymbol{\lambda}$ with, for some level $\ell \in \mathbb{N}_0$, $\||\boldsymbol{\lambda}|\|_\infty := \max_i |\lambda_i| \le \ell$, i.e., $\bar{\boldsymbol{\Lambda}}$ corresponds to a *full grid*, then, with Λ denoting the set of $\lambda \in \nabla$ with $|\lambda| \le \ell$, one has

$$R_{\bar{\boldsymbol{\Lambda}}}(A \otimes \cdots \otimes A)I_{\bar{\boldsymbol{\Lambda}}} = a(\Psi|_\Lambda, \Psi|_\Lambda) \otimes \cdots \otimes a(\Psi|_\Lambda, \Psi|_\Lambda).$$

We conclude that the application of $\mathbf{a}(\boldsymbol{\Psi}|_{\boldsymbol{\Lambda}}, \boldsymbol{\Psi}|_{\boldsymbol{\Lambda}})$ can be evaluated in $\mathcal{O}(n\#\boldsymbol{\Lambda}) = \mathcal{O}(\#\boldsymbol{\Lambda})$ operations.

Next, we consider $\boldsymbol{\Lambda}$ to correspond to a *sparse grid*, i.e., for some $\ell \in \mathbb{N}_0$, it is the set of all multi-indices $\boldsymbol{\lambda}$ with $\||\boldsymbol{\lambda}|\|_1 := \sum_i |\lambda_i| \le \ell$. For simplicity thinking here of $n = 2$, we write $R_{\boldsymbol{\Lambda}}(A \otimes A)I_{\boldsymbol{\Lambda}} = R_{\boldsymbol{\Lambda}}(A \otimes \mathrm{Id})(\mathrm{Id} \otimes A)I_{\boldsymbol{\Lambda}}$. In view of the subsequent application of $R_{\boldsymbol{\Lambda}}(A \otimes \mathrm{Id})$, we realize that we need the result of the application of $(\mathrm{Id} \otimes A)I_{\boldsymbol{\Lambda}}$ only on some finite subset of $\boldsymbol{\nabla}$. The generally smallest subset that can be selected is the corresponding full grid index set $\bar{\boldsymbol{\Lambda}}$ defined above, i.e., we have $R_{\boldsymbol{\Lambda}}(A \otimes A)I_{\boldsymbol{\Lambda}} = R_{\boldsymbol{\Lambda}}(A \otimes \mathrm{Id})I_{\bar{\boldsymbol{\Lambda}}}R_{\bar{\boldsymbol{\Lambda}}}(\mathrm{Id} \otimes A)I_{\boldsymbol{\Lambda}}$. Unfortunately, the applications of both $R_{\boldsymbol{\Lambda}}(A \otimes \mathrm{Id})I_{\bar{\boldsymbol{\Lambda}}}$ and $R_{\bar{\boldsymbol{\Lambda}}}(\mathrm{Id} \otimes A)I_{\boldsymbol{\Lambda}}$ require $\mathcal{O}(\#\bar{\boldsymbol{\Lambda}})$ operations, where, in the standard setting that $\#\{\lambda \in \nabla : |\lambda| = \ell\} \eqsim 2^\ell$, one has that $\#\boldsymbol{\Lambda} \eqsim \ell 2^\ell$ and $\#\bar{\boldsymbol{\Lambda}} \eqsim 4^\ell$.

(Here and in other places, with $C \eqsim D$ we mean that both $C \lesssim D$ and $C \gtrsim D$, with the first relation meaning that C can be bounded by some absolute multiple of D, and the second one being defined as $D \lesssim C$.)

To solve the above problem, we apply a key idea by Balder and Zenger in [1] for the hierarchical hat functions, that for more general functions was applied later in, e.g., [2–4]. We split A into the upper block matrix $U = [a(\psi_\lambda, \psi_\mu)]_{|\lambda| \le |\mu|}$ and the strictly lower block matrix $L = [a(\psi_\lambda, \psi_\mu)]_{|\lambda| > |\mu|}$. By definition of U, L, and $\boldsymbol{\Lambda}$, we have

$$(U \otimes \mathrm{Id})I_{\boldsymbol{\Lambda}} = I_{\boldsymbol{\Lambda}}R_{\boldsymbol{\Lambda}}(U \otimes \mathrm{Id})I_{\boldsymbol{\Lambda}}, \quad R_{\boldsymbol{\Lambda}}(L \otimes \mathrm{Id}) = R_{\boldsymbol{\Lambda}}(L \otimes \mathrm{Id})I_{\boldsymbol{\Lambda}}R_{\boldsymbol{\Lambda}},$$

from which we infer that

$$\begin{aligned} R_{\boldsymbol{\Lambda}}(A \otimes A)I_{\boldsymbol{\Lambda}} &= R_{\boldsymbol{\Lambda}}((U + L) \otimes A)I_{\boldsymbol{\Lambda}} \\ &= R_{\boldsymbol{\Lambda}}(L \otimes \mathrm{Id})(\mathrm{Id} \otimes A)I_{\boldsymbol{\Lambda}} + R_{\boldsymbol{\Lambda}}(U \otimes \mathrm{Id})(\mathrm{Id} \otimes A)I_{\boldsymbol{\Lambda}} \\ &= R_{\boldsymbol{\Lambda}}(L \otimes \mathrm{Id})(\mathrm{Id} \otimes A)I_{\boldsymbol{\Lambda}} + R_{\boldsymbol{\Lambda}}(\mathrm{Id} \otimes A)(U \otimes \mathrm{Id})I_{\boldsymbol{\Lambda}} \\ &= R_{\boldsymbol{\Lambda}}(L \otimes \mathrm{Id})I_{\boldsymbol{\Lambda}}\, R_{\boldsymbol{\Lambda}}(\mathrm{Id} \otimes A)I_{\boldsymbol{\Lambda}} + R_{\boldsymbol{\Lambda}}(\mathrm{Id} \otimes A)I_{\boldsymbol{\Lambda}}\, R_{\boldsymbol{\Lambda}}(U \otimes \mathrm{Id})I_{\boldsymbol{\Lambda}}. \end{aligned}$$

Since $\boldsymbol{\Lambda}$, "frozen" in either of its coordinates, is a collection of indices of all wavelets up to some level, $R_{\boldsymbol{\Lambda}}(\mathrm{Id} \otimes A)I_{\boldsymbol{\Lambda}}$, and similarly, $R_{\boldsymbol{\Lambda}}(L \otimes \mathrm{Id})I_{\boldsymbol{\Lambda}}$ and $R_{\boldsymbol{\Lambda}}(U \otimes \mathrm{Id})I_{\boldsymbol{\Lambda}}$, can be applied in $\mathcal{O}(\#\boldsymbol{\Lambda})$ operations, and so can $\mathbf{a}(\boldsymbol{\Psi}|_{\boldsymbol{\Lambda}}, \boldsymbol{\Psi}|_{\boldsymbol{\Lambda}})$.

Remark 1.1. Since, for general $n > 1$, in the above scheme two recursive calls for $\mathrm{Id} \otimes A \otimes \cdots \otimes A$ have to be made, one verifies that its complexity is $\mathcal{O}(2^n\#\boldsymbol{\Lambda})$. So in high dimensions, one should avoid multiplications with $R_{\boldsymbol{\Lambda}}(A_1 \otimes \cdots \otimes A_n)I_{\boldsymbol{\Lambda}}$ for more than one or a few A_i not being truly sparse. For, say, the Poisson problem, this can be realized by applying orthogonal wavelets (cf. [5,6]) or prewavelets (cf. [7,8]). This issue, however, is outside the scope of the current paper, and we will ignore the dependency of constants on the space dimension n.

The goal of this paper is to generalize the algorithm for the multiplication with $\mathbf{a}(\check{\boldsymbol{\Psi}}|_{\check{\boldsymbol{\Lambda}}}, \hat{\boldsymbol{\Psi}}|_{\hat{\boldsymbol{\Lambda}}})$ from [1] to the case of $\check{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\Lambda}}$ being *multi-trees*, and to prove that it requires only $\mathcal{O}(\#\check{\boldsymbol{\Lambda}} + \#\hat{\boldsymbol{\Lambda}})$ operations. We define a multi-index set $\boldsymbol{\Lambda}$ to be a multi-tree when "frozen" in any $n - 1$ coordinates, it is a *tree* in the remaining coordinate.

The application of this result lies in adaptive tensor product approximation methods, as adaptive sparse grid methods ([9] + references cited there), or adaptive tensor product wavelet Galerkin methods (e.g. [10]). It seems that multi-trees are the most general sets for which (1.1) is realizable (unless, by a special choice of the wavelets, the bi-infinite matrix $a(\check{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Psi}})$ as a whole is sparse, cf. [11]).

For $\check{\boldsymbol{\Psi}} = \hat{\boldsymbol{\Psi}}$ being a hierarchical basis and $\check{\boldsymbol{\Lambda}} = \hat{\boldsymbol{\Lambda}}$, similar results, although described more informally, can be found in [12,8,13]. For a hierarchical basis, our condition of $\check{\boldsymbol{\Lambda}} = \hat{\boldsymbol{\Lambda}}$ being a multi-tree is equal to the condition on this index set imposed in these references. The discussion in [8, Section 3.1.3] about a prewavelet basis learns that the generalization from the hierarchical basis to a general multi-level collection is not trivial.