Found Comput Math (2010) 10: 569–613
DOI 10.1007/s10208-010-9070-4

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

Scattering in Flatland: Efficient Representations via Wave Atoms

Laurent Demanet · Lexing Ying

Received: 23 May 2008 / Revised: 9 October 2009 / Accepted: 27 February 2010 /
Published online: 3 June 2010

Abstract This paper presents a numerical compression strategy for the boundary integral equation of acoustic scattering in two dimensions. These equations have oscillatory kernels that we represent in a basis of wave atoms, and compress by thresholding the small coefficients to zero.

This phenomenon was perhaps first observed in 1993 by Bradie, Coifman, and Grossman, in the context of local Fourier bases (Bradie et al. in Appl. Comput. Harmon. Anal. 1:94–99, 1993). Their results have since then been extended in various ways. The purpose of this paper is to bridge a theoretical gap and prove that a well-chosen fixed expansion, the non-standard wave atom form, provides a compression of the acoustic single- and double-layer potentials with wave number k as $O(k)$-by-$O(k)$ matrices with $C_{\epsilon\delta}k^{1+\delta}$ non-negligible entries, with $\delta > 0$ arbitrarily small, and ϵ the desired accuracy. The argument assumes smooth, separated, and not necessarily convex scatterers in two dimensions. The essential features of wave atoms that allow this result to be written as a theorem are a sharp time-frequency localization that wavelet packets do not obey, and a parabolic scaling (wavelength of the wave packet) $\sim$ (essential diameter)^2. Numerical experiments support the estimate and show that this wave atom representation may be of interest for applications where the same

Communicated by Albert Cohen.

L. Demanet (✉)
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
e-mail: demanet@gmail.com

L. Ying
Department of Mathematics, University of Texas at Austin, Austin, TX 78712, USA

scattering problem needs to be solved for many boundary conditions, for example, the computation of radar cross sections.

Keywords Fast algorithm · Wave propagation · Boundary integral equation · Computational harmonic analysis

Mathematics Subject Classification (2000) 65R20 · 65T99

1 Introduction

This paper is concerned with the sparse representation of the boundary integral operator of two-dimensional scattering problems. Let D be a bounded soft scatterer in $\mathbb{R}^2$ with a smooth boundary and $u^{\mathrm{inc}}(x)$ be the incoming wave field. The scattered field $u(x)$ satisfies the two-dimensional exterior Dirichlet problem of the Helmholtz equation:

$$\begin{aligned} &-\Delta u(x) - k^2 u(x) = 0 \quad \text{in } \mathbb{R}^d \setminus \bar{D}, \\ &u(x) = -u^{\mathrm{inc}}(x) \quad \text{for } x \in \partial D, \\ &\lim_{|x|\to\infty} |x|^{1/2} \left(\left(\frac{x}{|x|}, \nabla u(x) \right) - \mathrm{i}k u(x) \right) = 0. \end{aligned}$$

Typically, the higher the wave number k, the harder the computational problem. One attractive method for dealing with this equation is to reformulate it using a boundary integral equation for an unknown field $\phi(x)$ on ∂D:

$$\frac{1}{2}\phi(x) + \int_{\partial D} \left(\frac{\partial G(x,y)}{\partial n_y} - \mathrm{i}\eta G(x,y) \right) \phi(y)\,\mathrm{d}y = -u^{\mathrm{inc}}(x), \tag{1}$$

where n_y stands for the exterior normal direction of ∂D at the point y, and η is a coupling constant of order $O(k)$. This is the combined field boundary integral equation [18]. The kernels $G(x,y)$ and $\frac{\partial G(x,y)}{\partial n_y}$ are, respectively, the Green's function of the Helmholtz equation and its normal derivative, given by

$$G(x,y) = \frac{\mathrm{i}}{4} H_0^{(1)}\big(k\|x-y\|\big),$$

and

$$\frac{\partial G}{\partial n_y}(x,y) = \frac{\mathrm{i}k}{4} H_1^{(1)}\big(k\|x-y\|\big) \frac{x-y}{\|x-y\|} \cdot n_y.$$

Once $\phi(x)$ is obtained from solving the integral equation, the scattered field $u(x)$ at $x \in \mathbb{R}^2 \setminus \bar{D}$ can be evaluated as

$$u(x) = \int_{\partial D} \left(\frac{\partial G(x,y)}{\partial n_y} - \mathrm{i}\eta G(x,y) \right) \phi(y)\,\mathrm{d}y.$$

An important property of (1) from the computational point of view is that its condition number is often quite small and, as a result, one can advantageously solve (1) with an iterative algorithm like GMRES. At each step of the iterative solver, we need to apply the integral operator to a given function. Since the integral operator is dense, applying the operator directly is too expensive. In this paper, we address this issue by efficiently representing the operator as a sparse matrix in a system of wave atoms.

Local cosines or wavelet packets have already been proposed for this task with great practical success, see Sect. 1.5 for some references, but we believe that the following two reasons make a case for *wave atom frames*:

- The proposed construction is non-adaptive: wave atom frames of L^2 are not designed for a specific value of k, and no optimization algorithm is needed to find a provably good basis. To achieve this result, the essential property of wave atoms is a parabolic scaling that we discuss later.
- The choice of numerical realization for wave atoms follows some of the experience garnered throughout the 1990s in the study of local Fourier bases and wavelet packets. In particular, wave atoms offer a clean multiscale structure in the sense that they avoid the "frequency leakage" associated with wavelet packets defined from filterbanks. These aspects are discussed in [10].

Of course, any non-adaptive all-purpose numerical compression method is likely to lag in performance behind an adaptive strategy that would include at least the former in its scope; but this is no excuse for discarding their study. Proper insight about architectures and scalings is important for designing the solution around which, for instance, a library of bases should be deployed for a best basis search.

The main result of this paper says that the wave atom frame is in some sense near-optimal for representing the integral operator in (1) as a sparse matrix. Namely, full matrices would involve $O(k^2)$ elements but we show that $C_\epsilon k$ matrix elements suffice to represent G, and $C_{\epsilon\delta} k^{1+\delta}$ matrix elements suffice to represent $\partial G/\partial n_y$ to a given accuracy ϵ, for arbitrarily small $\delta > 0$. In the sequel we systematically write $O(k^{1+1/\infty})$ for the latter case. We believe that these bounds would not hold for wavelet packets, for instance, even if the best decomposition tree is chosen. In particular, wavelets would obviously not be suited for the job.

The potential implication of this result for scientific computing is discussed in Sect. 1.4 below, where it is explained that more ingredients are needed for obtaining fast algorithms.

1.1 Wave Atoms

Frames of wave atoms were introduced in [10] on the basis that they provide sparse representations of certain oscillatory patterns. As alluded to earlier, they are a special kind of oriented wavelet package which does not suffer from the frequency leaking associated to filterbanks, and which obeys the important parabolic scaling relation

$$\text{wavelength of the wave packet} \sim (\text{essential diameter})^2.$$

The wavelength of the wave packet is not to be confused with the wavelength $2\pi/k$ of the physical problem.

Let us recall the construction of wave atoms, and refer the reader to [10] for more details. In one dimension, wave atoms are an orthonormal basis indexed by the triple of integers $\lambda \equiv (j, m, n)$. The construction is in the frequency domain; our convention for the Fourier transform is

$$\hat{f}(\omega) = \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}x\cdot\omega} f(x)\,\mathrm{d}x, \qquad f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}x\cdot\omega} \hat{f}(\omega)\,\mathrm{d}\omega.$$

- First, $j \geq 0$ is a scale parameter that should be thought of as indexing dilations *of a factor 4*; in other words, one should consider a first partition of the positive frequency axis into intervals of the form $[c_1 2^{2j}, c_2 2^{2j+2}]$ (for some constants c_1, c_2 that will accommodate overlapping of basis functions). Choosing j such that frequency ω is proportional to 2^{2j} is in contrast with wavelet theory, where $\omega \sim 2^j$ over the positive frequency support of a wavelet.
- The parameter m with $c_1 2^j \leq m < c_2 2^{j+2}$ then indexes the further partitioning of each interval $[c_1 2^{2j}, c_2 2^{2j+2}]$ into $O(2^j)$ subintervals of size $O(2^j)$. More precisely, wave atoms are centered in frequency near $\pm\omega_\lambda$ where

  $$\omega_\lambda \approx \pi 2^j m, \qquad c_1 2^j \leq m < c_2 2^{j+2}$$

  and are compactly supported in the union of two intervals of length $2\pi \times 2^j$. The parabolic scaling is now apparent; the size of the support in frequency ($\sim 2^j$) is proportional to the square root of the offset from the origin ($\sim 2^{2j}$).
- The parameter $n \in \mathbb{Z}$ indexes translations. A wave atom is centered in space near

  $$x_\lambda = 2^{-j} n,$$

  and has essential support as narrow as the uncertainty principle allows, i.e., of length $O(2^{-j})$.

We define Ω to be the set of all admissible indices, i.e.,

$$\Omega = \big\{(j, m, n) : j \geq 0, c_1 2^j \leq m < c_2 2^{j+2}, n \in \mathbb{Z}\big\}.$$

Basis functions are then written

$$\varphi_\lambda(x) = 2^{j/2} \varphi_{(j,m)}\big(2^j x - n\big) \mathrm{e}^{\mathrm{i}2^j m x}, \quad \lambda \in \Omega, \tag{2}$$

where $\varphi_{(j,m)}$ depends weakly on j and m, and needs to be chosen adequately to form an orthobasis. The underlying delicate construction of the $\varphi_{(j,m)}$ is due to Villemoes [23] and summarized in [10].

In two dimensions, wave atoms are individually, but not collectively, formed as tensor products of the one-dimensional basis functions. The construction is "multiresolution" in the sense that there is only one dilation parameter; the indexing 5-tuple of integers is $\mu \equiv (j, \mathbf{m}, \mathbf{n})$ where $\mathbf{m} = (m_1, m_2)$ and $\mathbf{n} = (n_1, n_2)$. More precisely, at scale j, the valid values for $\mathbf{m} = (m_1, m_2)$ satisfy $0 \leq m_1, m_2 < c_2 2^{j+2}$ and $c_1 2^j \leq \max(m_1, m_2)$.

Wave atoms come as an orthonormal basis in two dimensions, but can be made fully directional—supported in a narrow cone in frequency with apex at the origin

[2]—at the expense of increasing the redundancy to two or four. The definition of such variants makes use of a unitary recombination involving Hilbert-transformed basis functions as in the definition of complex wavelet transforms, and is fully explained in [10].

None of the results of this paper would depend on the choice of variant; and for convenience we use the frame of wave atoms with redundancy four. With $\mathbf{x} = (x_1, x_2)$, the only property of wave atoms that we will need is the characterization

$$\varphi_\mu(\mathbf{x}) = 2^j \varphi_{(j,\mathbf{m})}\big(2^j x_1 - n_1, 2^j x_2 - n_2\big) \mathrm{e}^{\mathrm{i}2^j \mathbf{m}\cdot\mathbf{x}}, \tag{3}$$

where $\varphi_{(j,\mathbf{m})}$ is a C^∞ non-oscillatory bump that depends on j and $\mathbf{m}$, but in a non-essential manner, i.e.,

$$\big|\partial_{\mathbf{x}}^\alpha \varphi_{(j,\mathbf{m})}(\mathbf{x})\big| \le C_{\alpha,M}\big(1 + \|\mathbf{x}\|\big)^{-M}, \quad \forall M > 0, \tag{4}$$

with $C_{\alpha,M}$ independent of j and $\mathbf{m}$. In addition, each $\varphi_{(j,\mathbf{m})}$ is simply the tensor product of two corresponding bumps for the 1D transform.

Although they may not necessarily form an orthonormal basis, wave atoms still form a tight frame in the sense that expanding a function is an isometry from $L^2(\mathbb{R}^2)$ to $\ell_2(\mu)$,

$$\|f\|_2^2 = \sum_\mu \big|\langle f, \varphi_\mu\rangle\big|^2$$

which is equivalent to

$$f = \sum_\mu \langle f, \varphi_\mu\rangle \varphi_\mu. \tag{5}$$

The same properties hold in one dimension.

The closest analog to a "continuous wave atom transform" was introduced in the mathematical literature by Córdoba and Fefferman in [8]. Wave atoms can be compared to brushlets [20], but Villemoes's construction uses "local complex exponentials" instead of local cosines in frequency.

Discretized wave atoms are described in [9, 10]; they inherit the localization and tight-frame properties of their continuous counterpart. In particular, their bandlimited character confers an immediate control over the accuracy of computing inner products via quadrature. They come with fast FFT-based $O(N \log N)$ algorithms for both the forward and adjoint transforms (see [10] for details).

1.2 Operator Expansions

As functions can be analyzed and synthesized using coefficients, operators can also be expanded from matrix elements in a tight frame.

- The *standard form* of an operator A in the wave atom frame φ_λ is

$$A = \sum_{\lambda\in\Omega} \sum_{\lambda'\in\Omega} \varphi_\lambda A_{\lambda,\lambda'} \langle\cdot, \varphi_{\lambda'}\rangle,$$

where $A_{\lambda,\lambda'} = \langle A\varphi_{\lambda'}, \varphi_\lambda \rangle$. For each fixed λ, we define $S(\lambda)$ to be the set of all λ' such that the modulus of $A_{\lambda,\lambda'}$ is above a certain threshold. The sparse representation of A then takes the form

$$A \approx \sum_\lambda \varphi_\lambda \sum_{\lambda' \in S(\lambda)} A_{\lambda,\lambda'} \langle \cdot, \varphi_{\lambda'} \rangle.$$

In practice, the sums in λ and λ' are also truncated in scale to account for the finite number of samples N of the functions to which A is applied. The above equation naturally gives rise to an efficient method of applying the operator A to a given function f:

- Apply the forward wave atom transform to compute the coefficients $f_{\lambda'} := \langle f, \varphi_{\lambda'} \rangle$.
- For each λ, compute $g_\lambda := \sum_{\lambda' \in S(\lambda)} A_{\lambda,\lambda'} f_{\lambda'}$.
- Apply the adjoint wave atom transform to g_λ to synthesize Af, i.e., $Af \approx \sum_\lambda \varphi_\lambda g_\lambda$.

- The *non-standard form* of A in the two-dimensional frame φ_μ is the set of coefficients $A_\mu = \int_{\mathbb{R}^2} A(x_1, x_2) \overline{\varphi}_\mu(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2$, such that the distributional kernel of A is expanded as

$$A(x_1, x_2) = \sum_\mu A_\mu \varphi_\mu(x_1, x_2).$$

For a fixed threshold value, we define S to be the set of all μ such that A_μ is above the threshold in modulus. The sparse representation of A is now $A(x_1, x_2) \approx \sum_{\mu \in S} A_\mu \varphi_\mu(x_1, x_2)$. Applying A to a given function f efficiently using this expansion is more involved than the case of the standard form. For a fixed index $\mu = (j, \mathbf{m}, \mathbf{n})$ with $\mathbf{m} = (m_1, m_2)$ and $\mathbf{n} = (n_1, n_2)$, we define the two 1D wave atom indices:

$$\lambda_1^\mu = (j, m_1, n_1) \quad \text{and} \quad \lambda_2^\mu = (j, m_2, n_2). \tag{6}$$

Since the two-dimensional index $\mathbf{m} = (m_1, m_2)$ satisfies $0 \le m_1, m_2 < c_2 2^{j+2}$ and $c_1 2^j \le \max(m_1, m_2)$, the set of all possible choices for λ_1^μ and λ_2^μ are

$$0 \le m_1 < c_2 2^{j+2}, \quad \text{and} \quad 0 \le m_2 < c_2 2^{j+2}. \tag{7}$$

Some of the indices in (7) are not admissible, i.e., they are not part of the set of indices for the 1D wave atom transform, since if $\mu = (j, m, n)$ corresponds to a 1D wave atom it would need to satisfy $c_1 2^j \le m < c_2 2^{j+2}$. Non-admissible indices correspond to Gabor-type wave forms that partition the frequency domain uniformly and, hence, violate the parabolic scaling. We use Ω^e to denote this extended index set

$$\Omega^e = \big\{(j, m, n) : j \ge 0, 0 \le m < c_2 2^{j+2}, n \in \mathbb{Z}\big\}.$$

Again, such extended indices are needed because the 2D wave atom transform is not simply a tensor product of 1D transforms. The situation is the same for 2D MRA wavelets not being tensor products of two 1D wavelet bases.

The frame formed by the basis functions φ_λ with $\lambda \in \Omega^e$ is called the *extended* wave atom frame. For a given function f, the computation of all the coefficients $\langle f, \varphi_\lambda \rangle$ with $\lambda \in \Omega^e$ can be done easily by extending the existing forward wave atom transform to include the extra m indices ($0 \le m < c_1 2^j$) in Ω^e. The adjoint transform can be extended similarly and the resulting transform is called the *adjoint extended wave atom transform*. We note that these new transforms are not orthonormal any more since parts of the input functions are analyzed redundantly.

Using the notation in (6) and the tensor-product property of φ_μ, we have

$$A(x_1, x_2) \approx \sum_{\mu \in S} A_\mu \varphi_{\lambda_1^\mu}(x_1) \varphi_{\lambda_2^\mu}(x_2). \tag{8}$$

When A is applied to a given function f, we have

$$\begin{aligned} Af(x_1) &\approx \sum_{\mu \in S} A_\mu \varphi_{\lambda_1^\mu}(x_1) \left(\int \varphi_{\lambda_2^\mu}(x_2) f(x_2)\, \mathrm{d}x_2 \right) \\ &= \sum_{\lambda \in \Omega^e} \varphi_\lambda \sum_{\mu \in S : \lambda_1^\mu = \lambda} A_\mu \left(\int \varphi_{\lambda_2^\mu}(x_2) f(x_2)\, \mathrm{d}x_2 \right). \end{aligned}$$

Using the extended transforms, we can derive from the above equation a fast algorithm for applying A to f using the non-standard form:

- Apply the forward extended wave atom transform to compute the coefficients $f_{\lambda'} := \langle f, \varphi_{\lambda'} \rangle$ for all indices $\lambda' \in \Omega^e$.
- For each $\lambda \in \Omega^e$, compute $g_\lambda := \sum_{\mu \in S : \lambda_1^\mu = \lambda} A_\mu f_{\lambda_2^\mu}$.
- Apply the inverse extended wave atom transform to synthesize Af from g_λ, i.e., $Af \approx \sum_{\lambda \in \Omega^e} \varphi_\lambda g_\lambda$.

We would like to point out that non-standard expansions only exist for two-dimensional frames that have a tensor-product representation for each basis function, since the decomposition in (8) is essential for the derivation. For example, there is no known non-standard form representation for an operator in the tight frame of curvelets [6].

In what follows we focus exclusively on the non-standard form, because of its relative simplicity over the standard form. The isotropy of the envelope of φ_μ makes some of the stationary-phase arguments in the sequel simpler in our view. We can however not exclude at this point that the standard form may enjoy comparable sparsity properties as the non-standard form.

1.3 Sparsity of the Non-standard Wave Atom Matrix

In this section, we formulate the main result on sparsity of the non-standard wave atom matrix of the acoustic single- and double-layer potentials, in two dimensions.

The scatterer is a union of closed, non-intersecting C^∞ curves $\Omega = \bigcup_{\alpha=1}^n \Omega_\alpha$ embedded in $\mathbb{R}^2$. For each α, assume that $\mathbf{x}(t) : I_\alpha \mapsto \Omega_\alpha$ is a C^∞ periodic parametrization of Ω_α, and take $I_\alpha = [0, 1]$ for simplicity.

We assume the following mild *geometric regularity* condition on the scatterer: there exists $D > 0$ such that

$$\big\|\mathbf{x}(s) - \mathbf{x}(t)\big\| \geq D\big|\mathrm{e}^{2\pi \mathrm{i}s} - \mathrm{e}^{2\pi \mathrm{i}t}\big|, \tag{9}$$

essentially meaning that the curve Ω_α defining the scatterer cannot intersect itself. We write $d(s,t) \equiv |\mathrm{e}^{2\pi \mathrm{i}s} - \mathrm{e}^{2\pi \mathrm{i}t}|$ for the Euclidean distance on the unit circle.

When $s \in I_\alpha$, $t \in I_\beta$ with $1 \leq \alpha, \beta \leq n$, let

$$G_0(s,t) = \frac{\mathrm{i}}{4} H_0^{(1)}\big(k\big\|\mathbf{x}(s) - \mathbf{x}(t)\big\|\big)\big\|\dot{\mathbf{x}}(t)\big\|, \tag{10}$$

and

$$G_1(s,t) = \frac{\mathrm{i}k}{4} H_1^{(1)}\big(k\big\|\mathbf{x}(s) - \mathbf{x}(t)\big\|\big)\frac{\mathbf{x}(s) - \mathbf{x}(t)}{\|\mathbf{x}(s) - \mathbf{x}(t)\|} \cdot n_{\mathbf{x}(t)}\big\|\dot{\mathbf{x}}(t)\big\|. \tag{11}$$

The non-standard wave atom matrices of G_0 and G_1, restricted to a Cartesian product $I_\alpha \times I_\beta$ of intervals, are

$$K_\mu^0 = \langle G_0, \varphi_\mu \rangle, \qquad K_\mu^1 = \langle G_1, \varphi_\mu \rangle.$$

Our main result below concerns the existence of ϵ-*approximants* $\tilde{K}_\mu^0$ and $\tilde{K}_\mu^1$, corresponding to the restriction of μ to sets Λ_0 and Λ_1, i.e., with $a = 0, 1$,

$$\tilde{K}_\mu^a = \begin{cases} K_\mu^a & \text{if } \mu \in \Lambda_a; \\ 0 & \text{otherwise,} \end{cases}$$

and chosen by definition such that

$$\big\|K^a - \tilde{K}^a\big\|_{\ell_2(\mu)} \leq \epsilon, \quad a = 0, 1. \tag{12}$$

The ℓ_2 norm of a non-standard wave atom matrix is equivalent to a Hilbert–Schmidt norm for the corresponding operator, by the tight-frame property of wave atoms. This norm is of course stronger than the operator L^2-to-L^2 norm; and much stronger than the $\ell_\infty(\mu)$ norm used in [5].

In what follows the notation $A \lesssim B$ means $A \leq CB$ for some constant C that depends only on the non-essential parameters. Similarly, the notation $A \lesssim \epsilon^{-1/\infty}$ means $A \leq C_M \epsilon^{-1/M}$ for all $M > 0$. The constants may change from line to line.

Theorem 1 *Assume the scatterer is smooth and geometrically regular in the sense of (9). In the notations just introduced, there exist sets Λ_0 and Λ_1 that define ϵ-approximants of K_μ^0 and K_μ^1 respectively, and whose cardinality obeys*

$$|\Lambda_0| \leq C_M^0 \left[k\epsilon^{-1/M} + \left(\frac{1}{\epsilon}\right)^{2+1/M}\right], \tag{13}$$

$$|\Lambda_1| \leq C_M^1 \left[k^{1+1/M}\epsilon^{-1/M} + \left(\frac{k}{\epsilon}\right)^{2/3+1/M}\right], \tag{14}$$

for all $M > 0$, and where C_M^0, C_M^1 depend only on M and the geometry of the scatterers.

The terms proportional to k or $k^{1+1/\infty}$ are due to the oscillations when $s \neq t$, and the terms $\epsilon^{-2-1/\infty}$ and $(k/\epsilon)^{2/3+1/\infty}$ are due to the kernels' singularities on the diagonal $s = t$. While the growth rate in k for the oscillations term is smaller for G_0 than for G_1, the growth rate for the diagonal contribution is smaller for G_1 than for G_0 since

$$k^{2/3}\epsilon^{-2/3} = k^{1/3}k^{1/3}\epsilon^{-2/3} \leq \max\bigl(k, k, \epsilon^{-2}\bigr) \leq 3\bigl(k + \epsilon^{-2}\bigr).$$

Theorem 1 can also be formulated using *relative* errors instead of absolute errors. This viewpoint is important for considering the composed kernel

$$G_{(0,1)}(s,t) = G_1(s,t) - \mathrm{i}\eta G_0(s,t), \quad \eta \asymp k.$$

Call $K_\mu^{(0,1)}$ the non-standard wave atom matrix of $G_{(0,1)}$. In the following result, we quantify the number of terms needed to obtain the relative error estimate

$$\bigl\| K^a - \tilde{K}^a \bigr\|_{\ell_2(\mu)} \leq \epsilon \bigl\| K^a \bigr\|_{\ell_2(\mu)}, \quad a = 0, 1, \text{or } (0,1).$$

Corollary 2 *Let $\eta \asymp k$. In the assumptions and notations of Theorem 1, let Δ_0, Δ_1 and $\Delta_{(0,1)}$ be the sets of wave atom coefficients needed to represent the operators G_0, G_1, resp. $G_{(0,1)}$ up to relative accuracy ϵ. Then*

$$|\Delta_0| \leq C_M^0 \bigl[\bigl(k\epsilon^{-2}\bigr)^{1+1/M}\bigr], \tag{15}$$

$$|\Delta_1| \leq C_M^1 \bigl[k^{1+1/M}\epsilon^{-1/M} + \bigl(k\epsilon^{-2}\bigr)^{1/3+1/M}\bigr], \tag{16}$$

$$|\Delta_{(0,1)}| \leq C_M^{(0,1)} \bigl[\bigl(k\epsilon^{-2}\bigr)^{1+1/M}\bigr], \tag{17}$$

for all $M > 0$, and where the constants depend only on M and on the geometry of the scatterers.

We do not know if factors such as $k^{1/\infty}$ and $\epsilon^{-1/\infty}$ could be replaced by log factors.

The proofs of Theorem 1 and Corollary 2 occupy Sect. 2. The main ingredients are sparsity estimates in ℓ_p, stationary-phase considerations, vaguelette-type estimates adapted to wave atoms, and ℓ_2 correspondence scale-by-scale with wavelets.

It is interesting to note that the parabolic scaling of wave atoms is a necessary ingredient to obtain the right sparsity results. Any less oscillating basis functions (such as wavelets) would be too numerous to cover the support of the oscillatory kernel. Any more oscillating wave packets (such as Gabor) would require that too many frequencies are involved at any given location to recover the warped pattern to good accuracy. The parabolic scaling is also an essential ingredient in the dyadic–parabolic decomposition of Fourier integral operators, another kind of oscillatory integral [8].

1.4 Overview and Criticism of the Algorithm

Some numerical experiments that support the theory are presented in Sect. 3. The procedure followed for these experiments can be summarized as follows.

1. Evaluate the kernel on a Cartesian grid of N^2 points in (s,t) space, where N is proportional to k, the wave number. Form the non-standard representation of this kernel by taking the 2D wave atom transform of the sampled kernel—an $O(N^2 \log N)$ operation. Choose a threshold related to the eventual accuracy, and reduce to zero the wave atom coefficients below the threshold in magnitude—an $O(N^2)$ operation. The thresholded non-standard form is now available.
2. To apply the operator to a function, start by computing the extended 1D wave atom transform of the function, in $O(N \log N)$ operations. Compute the extended wave atom representation of the result of applying the operator, as discussed in Sect. 1.3. The complexity of this step ranges from $O(N^{1+1/\infty})$ to $O(N^2)$ depending on the data structure used for dealing with the wave atom matrix. Finally, compute an inverse extended wave atom transform in $O(N \log N)$ operations.

A word of caution on what has been achieved is in order.

First, this paper's numerical experiments are only meant to highlight the sparsity structure of the wave atom matrices. With a proper way to handle this sparsity at the level of the matrix's band structure, low-complexity $O(N^{1+1/\infty})$ algorithms could in principle be obtained for fast application of the operator itself. However, this paper only deals with full N-by-N matrices numerically, hence does not attempt to realize this speed-up in practice.

Second, even if wave atom matrices were properly stored as sparse banded matrices, there is the question of how to form this matrix other than by applying the wave atom transform directly to the kernel with $O(N^2)$ complexity. This question is not addressed here. It is less of a concern if the precomputation can be amortized over several applications of the matrix, such as when the scattering problem needs to be solved several times with different incoming waves $u^{\mathrm{inc}}(x)$. One important example is the computation of bistatic cross sections, where one needs to calculate the far-field patterns of scattered fields for all possible incoming plane waves. If the wave number k changes, however, there is at present no way to re-use previous computation.

Third, the numerical experiments presented here are only for "scattering in flatland" (in reference to the satirical novel by Abbott [1]), i.e., in two spatial dimensions. A solution similar in spirit in the three-dimensional case would involve more involved wave packet constructions. We also make no mathematical claim that sparsity would carry over for scatterers with corners, although we believe that similar estimates would hold. Singular scatterers would give rise to an interesting wavefront set for the kernel.

Finally, for inversion it is important to consider the condition number of the integral operator. It affects the speed of convergence of the scattering series. The conditioning question is mostly disjoint of that of realizing the operator numerically to some arbitrary precision. In our numerical experiments, compression of the direct operator neither helps nor hurts inversion in any substantial way.

1.5 Related Work

There has been a lot of work on sparsifying the integral operator of (1), or some variants of it, in appropriate bases. In [5], Bradie et al. showed that the operator becomes sparse in a local cosine basis. They proved that the number of coefficients with absolute value greater than any fixed ϵ is bounded by $O(k \log k)$ when the constant depends on ϵ. Notice that our result in Theorem 1 is stronger as the ℓ_2 norm is used instead in (12). In [3], Averbuch et al. extended the work in [5] by performing best basis search in the class of adaptive hierarchical local cosine bases.

Besides the local cosine transform, adaptive wavelet packets have been used to sparsify the integral operator as well. Deng and Ling [11] applied the best basis algorithm to the integral operator to choose the right one-dimensional wavelet packet basis. Golik [15] independently proposed to apply the best basis algorithm on the right-hand side of the integral equation (1). Shortly afterwards, Deng and Ling [12] gave similar results by using a predefined wavelet packet basis that refines the frequency domain near k. All of these approaches work with the standard form expansion of the integral operator. Recently in [16], Huybrechs and Vandewalle used the best basis algorithm for two-dimensional wavelet packets to construct a non-standard sparse expansion of the integral operator. In all of these results, the numbers of non-negligible coefficients in the expansions were reported to scale like $O(k^{4/3})$. However, our result shows that, by using the non-standard form based on wave atoms, the number of significant coefficients scales like $O(k^{1+1/\infty})$.

Most of the approaches on sparsifying (1) in well-chosen bases require the construction of the full integral operator. Since this step itself takes $O(k^2)$ operations, it poses a computational difficulty for large k values. In [4], Beylkin et al. proposed a solution to the related problem of sparsifying the boundary integral operator of the Laplace equation. They successfully avoided the construction of the full integral operator by predicting the location of the large coefficients and applying a special one-point quadrature rule to compute the coefficients. The corresponding solution for the integral operator of the Helmholtz equation is still missing.

There has been a different class of methods, initiated by Rokhlin in [21, 22], that requires no construction of the integral operator and takes $O(k \log k)$ operations in 2D to apply the integral operator. A common feature of these methods [7, 13, 14, 21, 22] is that they partition the spatial domain hierarchically with a tree structure and compute the interaction between the tree nodes in a multiscale fashion: Whenever two nodes of the tree are well separated, the interaction (of the integral operator) between them is accelerated either by Fourier transform-type techniques [7, 21, 22] or by directional low-rank representations [13, 14].

A criticism of the methods in [7, 13, 14, 21, 22] is that the constant in front of the complexity $O(k \log k)$ is often quite high. On the other hand, since the FFT-based wave atom transforms are extremely efficient, applying the operator in the wave atom frame has a very small constant once the non-standard sparse representation is constructed. Therefore, for applications where one needs to solve the same Helmholtz equation with many different right-hand sides, the current approach based on the wave atom basis can potentially offer a competitive alternative. As mentioned earlier, one important example is the computation of the radar cross section.

2 Sparsity Analysis

This section contains the proof of Theorem 1. The overarching strategy is to reduce the ℓ_2 approximation problem to an estimate of ℓ_p sparsity through a basic result of approximation theory, the direct "Jackson" estimate

$$\left\| K_\mu - \tilde{K}_\mu \right\|_2 \leq C |\Lambda|^{\frac{1}{2}-\frac{1}{p}} \|K_\mu\|_p,$$

where $\|K_\mu\|_p^p = \sum_\mu |K_\mu|^p$. Here $\tilde{K}_\mu$ refers to the approximation of K_μ where only the $|\Lambda|$ largest terms in magnitude are kept, and the others put to zero. The inequality is valid for all values of $0 < p < 2$ for which $\|K_\mu\|_p$ is finite. For a proof, see [19], p. 390.

If the ℓ_2 error is to be made less than ϵ, it is enough to have K_μ in some ℓ_p space, $0 < p < 2$, and take the number of terms defining $\tilde{K}_\mu$ to be

$$|\Lambda| \geq C_p \epsilon^{\frac{2p}{p-2}} \|K_\mu\|_p^{\frac{2p}{2-p}} \tag{18}$$

for some adequate $C_p > 0$. The sequence K_μ will be split into several fragments that will be studied independently. For each of these fragments F in μ space, the inequality (18) will be complemented by an estimate of the form $\|K_\mu\|_{\ell_p(F)} \leq C_p k^{q(p)}$, for all $p > p_0$. Three scenarios will occur in the sequel:

- If
$$p_0 = 0 \quad \text{and} \quad q(p) = \frac{1}{p} - \frac{1}{2}, \tag{19}$$
then $|F| \lesssim k\epsilon^{-1/\infty}$, which is the first term in (13).
- If
$$p_0 = 1 \quad \text{and} \quad q(p) = 0, \tag{20}$$
then $|F| \lesssim \epsilon^{-2-1/\infty}$, which is the second term in (13).
- If
$$p_0 = \frac{1}{2} \quad \text{and} \quad q(p) = \frac{1}{p} - 1 + \delta, \tag{21}$$
for arbitrarily small $\delta > 0$, then $|F| \lesssim (k/\epsilon)^{2/3+1/\infty}$, which is the second term in (14).

The problem is therefore reduced to identifying contributions in the sequence K_μ that obey either one of the three estimates above. In what follows we focus on the kernel $K = G_0$. We mention in Sect. 2.9 how the proof needs to be modified to treat the kernel G_1.

2.1 Smoothness of Hankel Functions

Bessel and Hankel functions have well-known asymptotic expansions near the origin and near infinity. That these asymptotic behaviors also determine smoothness in a sharp way over the whole half-line is perhaps less well-known, so we formulate these results as lemmas that we prove in the appendix.

Lemma 1 *For all integers $m \geq 0$ and $n \geq 0$ there exists $C_{m,n} > 0$ such that for all $k > 0$,*

$$\left|\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^m \left[\mathrm{e}^{-\mathrm{i}kx} H_n^{(1)}(kx)\right]\right| \leq \begin{cases} C_{mn}(kx)^{-1/2} x^{-m} & \text{if } kx \geq 1; \\ C_{mn}(kx)^{-n} x^{-m} & \text{if } 0 < kx < 1 \text{ and } m + n > 0; \\ C(1 + |\log kx|) & \text{if } 0 < kx < 1 \text{ and } m = n = 0. \end{cases} \tag{22}$$

The same results hold if 1 is replaced by any number $c > 0$ in $kx < 1$ vs. $kx \geq 1$.

The point of (22) is that $C_{m,n}$ is independent of k. Slightly more regularity can be obtained near the origin when multiplying with an adequate power of x, as the following lemma shows in the case of $H_1^{(1)}$.

Lemma 2 *For every integer $m \geq 0$ there exists $C_m > 0$ such that, for $0 < x \leq 1$,*

$$\left|\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^m \left[x H_1^{(1)}(x)\right]\right| \leq \begin{cases} C_m & \text{if } m = 0, 1; \\ C_2(1 + |\log x|) & \text{if } m = 2; \\ C_m x^{2-m} & \text{if } m > 2. \end{cases} \tag{23}$$

Finally, we will need the following lower bound.

Lemma 3 *For each $n \geq 0$, there exist $c_n > 0$ and $C_n > 0$ such that, when $x > c_n$,*

$$\left|H_n^{(1)}(x)\right| \geq C_n x^{-1/2}.$$

2.2 Dyadic Partitioning

Consider K as in Theorem 1, with $s \in I_\alpha$ and $t \in I_\beta$. If $\alpha = \beta$, K presents a singularity on its diagonal, whereas if $\alpha \neq \beta$ it presents no such singularity. The case $\alpha = \beta$ is representative and is treated in the sequel without loss of generality.

In this section we assume, as we have above, that $I_\alpha = [0, 1]$. The first step of the proof is to partition the periodized square $I_\alpha \times I_\alpha$ at each scale j, into dyadic squares denoted

$$Q = \left[2^{-j} q_1, 2^{-j}(q_1 + 1)\right] \times \left[2^{-j} q_2, 2^{-j}(q_2 + 1)\right], \quad q_1, q_2 \in \mathbb{Z}^+.$$

We define w_Q a window localized near Q through

$$w_Q(s, t) = w\left(2^j s - q_1, 2^j t - q_2\right),$$

where w is compactly supported on $[-1, 2]^2$ and of class C^∞. As a result w_Q is compactly supported inside

$$3Q \equiv \left[2^{-j}(q_1 - 1), 2^{-j}(q_1 + 2)\right] \times \left[2^{-j}(q_2 - 1), 2^{-j}(q_2 + 2)\right].$$

We also write $x_Q = (2^{-j} q_1, 2^{-j} q_2)$ for the bottom-left corner of Q, not to be confused with $\mathbf{x}$, which is in physical space.

Denote by $\mathcal{Q}_j$ the set of dyadic squares at scale j; we assume that w is chosen so that we have the scale-by-scale partition of unity property

$$\sum_{Q\in\mathcal{Q}_j} w_Q = 1.$$

The kernel is now analyzed at each scale j as $K = \sum_{Q\in\mathcal{Q}_j} K_Q$, where $K_Q = w_Q K$. We take the scale j of the dyadic partitioning to match the scale j in the wave atom expansion; namely if $\mu = (j, \mathbf{m}, \mathbf{n})$, then

$$K_\mu = \langle K, \varphi_\mu\rangle = \sum_{Q\in\mathcal{Q}_j} \langle K_Q, \varphi_\mu\rangle.$$

When $0 < p \le 1$, an estimate on the total ℓ_p norm can then be obtained from the p-triangle inequality, as follows:

$$\sum_j \sum_{\mathbf{m}} \sum_{\mathbf{n}} |K_{j,\mathbf{m},\mathbf{n}}|^p \le \sum_j \sum_{Q\in\mathcal{Q}_j} \sum_{\mathbf{m}} \sum_{\mathbf{n}} \left|\langle K_Q, \varphi_{j,\mathbf{m},\mathbf{n}}\rangle\right|^p. \tag{24}$$

When $p \ge 1$, then the regular triangle inequality will be invoked instead, for instance as in

$$\left(\sum_j \sum_{\mathbf{m}} \sum_{\mathbf{n}} |K_{j,\mathbf{m},\mathbf{n}}|^p\right)^{1/p} \le \sum_j \sum_{Q\in\mathcal{Q}_j} \left(\sum_{\mathbf{m}} \sum_{\mathbf{n}} \left|\langle K_Q, \varphi_{j,\mathbf{m},\mathbf{n}}\rangle\right|^p\right)^{1/p}. \tag{25}$$

The rationale for introducing a partitioning into dyadic squares is the technical fact that wave atoms are not built compactly supported in space. The windows w_Q allow to cleanly separate different regions of the parameter patch in which the kernel K oscillates with different local wave vectors. Note also that the dyadic partitioning is a mathematical tool for the proof of Theorem 1, and is not part of the construction of the wave atom transform.

The proof's architecture is summarized in the table at the end of this section. Dyadic squares are first classified according to their location with respect to the diagonal $s = t$, where K is singular.

1. *Diagonal squares.* Dyadic squares will be considered "diagonal squares" as soon as the distance from their center to the diagonal $s = t$ is less than $1/k$. We call the locus $S = \{d(s,t) \le 1/k\}$ the diagonal strip. Scale-by-scale, the condition on the square's centers reads $d(2^{-j}q_1, 2^{-j}q_2) \le 3\max(2^{-j}, \frac{1}{k})$. (We need to use the circle distance d since q_1 and q_2 are defined modulo 2^j.) There are $O(2^j \max(1, 2^j/k))$ such diagonal squares at scale j. They correspond to the case $kx \lesssim 1$ in Lemma 1.
2. *Non-diagonal squares.* When $d(2^{-j}q_1, 2^{-j}q_2) > 3\max(2^{-j}, 1/k)$, we say the square is non-diagonal. In those squares, the kernel K_Q is C^∞ but oscillatory. There are $O(2^{2j})$ such non-diagonal squares at scale j. They correspond to the case $kx \gtrsim 1$ in Lemma 1.

Non-diagonal squares are further subdivided into *near-field* and *far-field* squares, depending on their distance to the diagonal. The threshold is at $d(s,t) \le C$ for some constant C, determined as follows. Lemma 1 identifies the argument of the Hankel function as a *phase*. For the kernel K, this phase is $k\phi(s,t)$ where $\phi(s,t) = \|\mathbf{x}(s) - \mathbf{x}(t)\|$. The constant C mentioned earlier is chosen such that ϕ cannot be stationary in the near-field; however, there may be stationary points in the far-field. This distinction is important for the ℓ_p summation estimate in Sect. 2.5. A counting argument for the number of dyadic square where the phase ϕ is near-stationary is provided in Sect. 2.3.

Diagonal squares need to be further partitioned, or need further classifying, depending on the scale j. When the scale is large (j small), a diagonal dyadic square may not be contained inside the diagonal strip S; the triangular portions that extend outside the strip are smoothly cut out and give rise to the *off-strip* contribution. When the scale is small (j large), a square in the diagonal strip may intersect the diagonal $s = t$, or may not. The former case gives rise to the *singular on-strip* contribution, and the latter case is the *regular on-strip* contribution. More details on this subdivision are given at the beginning of Sect. 2.6. It is important to make a distinction between these contributions as they give rise to very different decay estimates for the wave atom matrix elements.

Non-diagonal squares		Diagonal squares		
Near-field	Far-field	Off-strip	On-strip	
			Singular	Regular

2.3 Geometry of Stationary-Phase Points

The phase $\phi(s,t) = \|\mathbf{x}(s) - \mathbf{x}(t)\|$ mentioned earlier generates typical oscillations as long as $\nabla\phi$ has large magnitude. On the other hand, we recognize $\nabla\phi = 0$ as being the "stationary point set" for the kernel considered. In this section, we argue that the locus of near-critical (or near-stationary) points of ϕ necessarily has small measure. The following lemma makes this heuristic precise in terms of the *scale defect* j'.

Lemma 4 *Let $\phi(s,t) = \|\mathbf{x}(s) - \mathbf{x}(t)\|$ for s,t in some I_α. For $j' \ge 0$, let*

$$\mathcal{K}_j(j') = \{(q_1,q_2) : \|\nabla\phi(2^{-j}q_1, 2^{-j}q_2)\|_\infty \le 2^{-j'}\}.$$

Then there exists $C > 0$ such that the cardinality of $\mathcal{K}_j(j')$ obeys

$$|\mathcal{K}_j(j')| \le C 2^{j+(j-j')_+}$$

where $(x)_+ = x$ if $x \ge 0$, and zero otherwise.

Proof Let $\mathbf{r} = (\mathbf{x}(s) - \mathbf{x}(t))/\|\mathbf{x}(s) - \mathbf{x}(t)\|$; the gradient of the phase is $\nabla\phi(s,t) = (\dot{\mathbf{x}}(s)\cdot r, -\dot{\mathbf{x}}(t)\cdot\mathbf{r})$. The condition $\|\nabla\phi(s,t)\|_\infty \le 2^{-j'}$, i.e.

$$|\dot{\mathbf{x}}(s)\cdot\mathbf{r}| \le 2^{-j'} \quad \text{and} \quad |\dot{\mathbf{x}}(t)\cdot\mathbf{r}| \le 2^{-j'},$$

is for large j' an almost-perpendicularity condition between tangent vectors to the curve Ω_α and the chord joining $\mathbf{x}(s)$ and $\mathbf{x}(t)$.

Now fix $s = 2^{-j}q_1$, and let $\mathbf{n}(s)$ be either normal vector to Ω_α at $\mathbf{x}(s)$. Let θ be the angle between $\mathbf{r}$ and $\mathbf{n}(s)$, such that $|\dot{\mathbf{x}}(s)\cdot\mathbf{r}| = \|\dot{\mathbf{x}}(s)\||\sin\theta|$. Since the parametrization is non-degenerate, the first condition $|\dot{\mathbf{x}}(s)\cdot\mathbf{r}| \le 2^{-j'}$ implies $\theta \le C2^{-j'}$ for some adequately large $C > 0$.

Consider therefore a cone Γ with apex at $\mathbf{x}(s)$, axis $\mathbf{n}(s)$, and opening $\theta \le C2^{-j'}$. The second condition $|\dot{\mathbf{x}}(t)\cdot\mathbf{r}| \le 2^{-j'}$ is satisfied only if the curve Ω_α intersects a chord inside the cone at a near-right angle, and as a consequence, every chord inside the cone at a near-right angle, differing from $\pi/2$ by $O(2^{-j'})$. Because Ω_α has finite length, bounded curvature, and obeys the geometric regularity property (9), there can only be a finite number of such intersections. The total length of $\Omega_\alpha \cap \Gamma$ is therefore $O(2^{-j'})$.

Since the points $\mathbf{x}(2^{-j}q_2)$ are a distance $C2^{-j}$ apart from each other, there are at most $O(\max(1, 2^{j-j'}))$ points indexed by q_2 that obey the two almost-orthogonality conditions, which can be written as $O(2^{(j-j')_+})$. Since q_1 takes on $O(2^j)$ values, the total number of couples (q_1, q_2) obeying the conditions is $O(2^j 2^{(j-j')_+})$. □

We will also need the observation that near-stationary-phase points can only occur far away from the diagonal.

Lemma 5 *As before, let* $d(s,t) = |e^{2\pi i s} - e^{2\pi i t}|$. *There exist two constants* $C_1, C_2 > 0$ *such that, if* $d(s,t) \le C_1$, *then*

$$\big\|\nabla\phi(s,t)\big\|_\infty \ge C_2.$$

Proof As previously,

$$\big\|\nabla\phi(s,t)\big\|_\infty = \min\big(\big|\dot{\mathbf{x}}(s)\cdot\mathbf{r}\big|, \big|\dot{\mathbf{x}}(t)\cdot\mathbf{r}\big|\big),$$

and we write $|\dot{\mathbf{x}}(s)\cdot\mathbf{r}|$ as $\|\dot{\mathbf{x}}(s)\||\cos(\theta_s)|$, where θ_s is the angle between the chord $(\mathbf{x}(s), \mathbf{x}(t))$ and the tangent vector $\dot{\mathbf{x}}(s)$. This angle obeys $|\theta_s| \lesssim d(s,t)|$, hence the cosine factor is greater than 1/2 as long as $d(s,t) \le C_1$ for some adequate C_1. The factor $\|\dot{\mathbf{x}}(s)\|$ is also bounded away from zero by regularity of the parametrization. The same argument can be made for $|\dot{\mathbf{x}}(t)\cdot\mathbf{r}|$. □

2.4 Non-diagonal Kernel Fragments: Decay of Individual Coefficients

The intuition for this section is that wave atom coefficients are small whenever their wavenumber differs from the local wave number of the kernel. It is an integration by parts argument, and it is not entirely trivial for two reasons: (1) the length scale of the decay in coefficient space must be chosen carefully as a function of k (parameter β below), and (2) Bessel functions have a leading-order $1/\sqrt{x}$ decay that must be preserved throughout the differentiations.

Within non-diagonal squares, $d(s,t) \gtrsim 1/k$ and $k\phi(s,t) \gtrsim 1$, so Lemma 1 asserts that K can be written as

$$K(s,t) = e^{ik\phi(s,t)} a\big(k\phi(s,t), s, t\big),$$

where $\phi(s,t) = \|\mathbf{x}(s) - \mathbf{x}(t)\|$ and the dependence of a on k is mild in comparison to that of $\mathrm{e}^{\mathrm{i}k\phi}$;

$$\left|\frac{\mathrm{d}^n}{\mathrm{d}\phi^n} a\big(k\phi(s,t),s,t\big)\right| \le C_n \frac{1}{\sqrt{k\phi(s,t)}} \phi(s,t)^{-n}.$$

The presence of additional arguments s and t is needed to account for factors such as the Jacobian $\|\mathbf{x}'(t)\|$; all the derivatives of these factors are $O(1)$ by assumption. Therefore, the chain rule yields

$$\left|\frac{\mathrm{d}^{\alpha_1}}{\mathrm{d}s^{\alpha_1}} \frac{\mathrm{d}^{\alpha_2}}{\mathrm{d}t^{\alpha_2}} a\big(k\phi(s,t),s,t\big)\right| \le C_\alpha \frac{1}{\sqrt{k\phi(s,t)}} \phi(s,t)^{-|\alpha|}, \quad \phi(s,t) \lesssim 1. \tag{26}$$

Fix $j > 0$ and $Q \in \mathcal{Q}_j$. We seek a good bound on

$$\begin{aligned}\langle K_Q, \varphi_{j,\mathbf{m},\mathbf{n}}\rangle = \int_{3Q} & w_Q(s,t) a\big(k\phi(s,t),s,t\big) \mathrm{e}^{\mathrm{i}k\phi(s,t)} \mathrm{e}^{-\mathrm{i}2^j \mathbf{m}\cdot(s,t)} \\ & \times 2^j \varphi_{(j,\mathbf{m})}\big(2^j s - n_1, 2^j t - n_2\big)\,\mathrm{d}s\,\mathrm{d}t,\end{aligned}$$

where $\varphi_{(j,\mathbf{m})}$ has been introduced in (3). Without loss of generality, we perform a translation to choose the coordinates s and t such that $x_Q = 0$ and $w_Q(s,t) = w(2^j s, 2^j t)$.

A first bound estimating the decay in $\mathbf{n}$ can be obtained by using (1) the almost-exponential decay (4) for $\varphi_{(j,\mathbf{m})}$, (2) the estimate $\|w_Q\|_{L^1} \lesssim 2^{-2j}$ that follows from $|3Q| \lesssim 2^{-2j}$, and (3) an L^∞ bound for the rest of the integrand, disregarding the oscillations. The result is

$$\big|\langle K_Q, \varphi_{j,\mathbf{m},\mathbf{n}}\rangle\big| \le C_M 2^{-j} \sup_{(s,t)\in 3Q} \big[\big(k\phi(s,t)\big)^{-1/2}\big]\big(1 + \|\mathbf{n}\|\big)^{-M}, \quad \forall M > 0.$$

The size of the first-order Taylor remainder of $k\phi(s,t)$ over $3Q$ is $O(2^j)$ times smaller that the value of $k\phi(x_Q)$ itself, so we may evaluate ϕ at x_Q at the expense of a multiplicative constant in the estimate. We get

$$\big|\langle K_Q, \varphi_{j,\mathbf{m},\mathbf{n}}\rangle\big| \le C_M 2^{-j} \big(k\phi(x_Q)\big)^{-1/2}\big(1 + \|\mathbf{n}\|\big)^{-M}, \quad \forall M > 0. \tag{27}$$

Capturing the decay in $\mathbf{m}$, however, requires integrations by parts. Heuristically, the objective is to show that the wave atom coefficients decay almost exponentially in $\mathbf{m}$, with a length scale of 1 in all directions (in units of $\mathbf{m}$), independently of j—at least in the representative case $j \simeq \frac{1}{2}\log_2 k$. To this end let us introduce the self-adjoint differential operator

$$L = \frac{I - \beta\Delta_{(s,t)} - \mathrm{i}\beta k(\Delta\phi(s,t))}{1 + \beta\|k\nabla\phi(s,t) - 2^j\mathbf{m}\|^2},$$

with

$$\beta = \frac{1}{\max(2^{-2j}k^2, 2^{2j})}.$$

We see that L leaves the exponential $\exp i[k\nabla\phi(s,t) - 2^j\mathbf{m}\cdot(s,t)]$ unchanged, hence we introduce M copies of L, and integrate by parts in s and t to pass the differentiations to the non-oscillatory factors. The scaling parameter β has been chosen such that the repeated action of L on the rest of the integrand introduces powers of $1/(1+\beta\|k\nabla\phi(s,t)-2^j\mathbf{m}\|^2)$, but otherwise only worsens the bound by a constant independent of $\mu = (j,\mathbf{m},\mathbf{n})$. Indeed, $\beta \le 2^{-2j}$, and

- The action of each derivative on w_Q or $\varphi_{(j,\mathbf{m})}$ produces a factor 2^j balanced by $\sqrt{\beta}$.
- The action of each derivative on a produces a factor $1/\phi(s,t)$, which by (9) is comparable to $1/d(s,t)$. Since we are in the presence of non-diagonal squares, $1/d(s,t) \lesssim \min(2^j,k) \le 2^j$. Again, each derivative produces a factor 2^j, which is balanced by $\sqrt{\beta}$. Note that the leading factor $1/\sqrt{k\phi}$ in the bound (26) is harmless since it is carried through the differentiations.

It is then tedious but straightforward to combine these observations and conclude that, for all $M > 0$,

$$\begin{aligned}&\left|L^M\left[w\left(2^j s, 2^j t\right)a\left(k\phi(s,t),s,t\right)\varphi_{(j,\mathbf{m})}\left(2^j s - n_1, 2^j t - n_2\right)\right]\right| \\ &\quad\le C_M \frac{1}{\sqrt{k\phi(s,t)}} \frac{1}{(1+\beta\|k\nabla\phi(s,t)-2^j\mathbf{m}\|^2)^M}.\end{aligned}$$

Since L is a differential operator, the support of the integrand remains $3Q$ regardless of M, hence we still get a factor $|3Q| \sim 2^{-2j}$ from the integral over s and t. With the L^2 normalization factor 2^j coming from (3), the overall dependence on scale is 2^{-j}. The resulting bound is

$$\begin{aligned}\left|\langle K_Q, \varphi_{j,\mathbf{m},\mathbf{n}}\rangle\right| \le C_M 2^{-j} \sup_{(s,t)\in 3Q} \Big[&\left(k\phi(s,t)\right)^{-1/2} \\ &\times \left(1+\beta\left\|k\nabla\phi(s,t)-2^j\mathbf{m}\right\|^2\right)^{-M}\Big], \quad \forall M > 0. \end{aligned} \tag{28}$$

The second factor inside the square brackets can be written as

$$\left(1+\left\|\frac{k2^{-j}\nabla\phi(s,t)-\mathbf{m}}{2^{-j}\beta^{-1/2}}\right\|^2\right)^{-M},$$

showing that in m-space, it is a fast-decaying bump centered at $k2^{-j}\nabla\phi(s,t)$ and of characteristic width $2^{-j}\beta^{-1/2}$. Over the set $3Q$, we have the estimate $|k2^{-j}\nabla\phi(s,t) - k2^{-j}\nabla\phi(x_Q)| = O(k2^{-2j})$. The quantity $k2^{-2j}$ is in all cases less than the length scale $2^{-j}\beta^{-1/2}$ (which is why we could not simply have taken $\beta = 2^{-2j}$), so we may replace $\nabla\phi(s,t)$ by $\nabla\phi(x_Q)$ in the expression of the bump, at the expense of a multiplicative constant depending only on M.

We have also seen earlier that $(k\phi(s,t))^{-1/2}$ can safely be replaced by $(k\phi(x_Q))^{-1/2}$ in the region $d(s,t) \gtrsim 1$, at the expense of another multiplicative constant. With these observations, we can take the geometric mean of (27) and (28) and obtain the central bound

$$\left|\langle K_Q, \varphi_{j,\mathbf{m},\mathbf{n}}\rangle\right| \le C_M 2^{-j}\left(k\phi(x_Q)\right)^{-1/2}\left(1+\beta\left\|k\nabla\phi(x_Q)-2^j\mathbf{m}\right\|^2\right)^{-M}\left(1+\|\mathbf{n}\|\right)^{-M}, \tag{29}$$

for all $M > 0$.

2.5 Non-diagonal Kernel Fragments: ℓ_p Summation

The expression just obtained can be used to show ℓ_p summability and verify proper growth as a function of k. Only the case $p \le 1$ is interesting and treated in this section. We tackle the different sums in the right-hand side of (24) in the order as written, from right to left.

- The sum over $\mathbf{n}$ is readily seen to contribute a multiplicative constant independent of the other parameters j, Q, and $\mathbf{m}$.
- Consider the sum over $\mathbf{m}$, and pull out the factor $2^{-jp}(k\phi(x_Q))^{-p/2}$. (We will not worry about this factor until we treat the sum over Q.) The range of values for $\mathbf{m}$ is an annulus $\|\mathbf{m}\|_\infty \asymp 2^j$, so we can compare the sum over $\mathbf{m}$ to the integral

$$\int_{C_j} \bigl(1+\beta\bigl\|k\nabla\phi(x_Q)-2^j\mathbf{x}\bigr\|^2\bigr)^{-Mp}\,\mathrm{d}\mathbf{x},$$

where $C_j = \{\mathbf{x}\in\mathbb{R}^2 : C_1 2^j \le \|\mathbf{x}\|_\infty \le C_2 2^j\}$ for some $C_1, C_2 > 0$. In what follows, take M sufficiently large so that, say, $Mp \ge 5$. Two cases need to be considered, corresponding to $2^{2j} \le k$ (large scales), and $2^{2j} > k$ (small scales).

– If $2^{2j} \le k$, then $\beta = 2^{2j}k^{-2}$. It will be sufficient to consider only the upper bound for $\|x\|_\infty$, whence we have the bound

$$\int_{\|\mathbf{x}\|_\infty\le C2^j}\left(1+\left\|\frac{2^{-j}k\nabla\phi(x_Q)-\mathbf{x}}{2^{-2j}k}\right\|^2\right)^{-Mp}\mathrm{d}\mathbf{x}.$$

With Lemma 4 in mind, we introduce the *scale defect* j'_Q as the unique integer such that

$$\frac{1}{2}2^{-j'_Q} < \bigl\|\nabla\phi(x_Q)\bigr\| \le 2^{-j'_Q}.$$

The integrand is a bump that essentially lies outside of the region of integration as soon as $k2^{-(j+j'_Q)} \gtrsim 2^j$.

More precisely, observe that

$$\begin{aligned}&\sup_{\mathbf{x}:\|\mathbf{x}\|_\infty\le C2^j}\left(1+\left\|\frac{2^{-j}k\nabla\phi(x_Q)-\mathbf{x}}{2^{-2j}k}\right\|^2\right)^{-Mp}\\&\quad\le C\bigl(1+2^j\bigl(2^{-j'_Q}-C'k^{-1}2^{2j}\bigr)_+\bigr)^{-2Mp}\end{aligned}$$

hence the integral is bounded by a first expression,

$$C2^{2j}\bigl(1+2^j\bigl(2^{-j'_Q}-Ck^{-1}2^{2j}\bigr)_+\bigr)^{-2Mp}. \tag{30}$$

A second bound can be obtained by letting $\mathbf{x}' = \mathbf{x} - 2^{-j}k\nabla\phi(x_Q)$ and extending the region of integration to the complement of a square in $\mathbf{x}'$, of the form

$$\bigl\|\mathbf{x}'\bigr\|_\infty \ge s_{jQ} = \left(\frac{1}{2}k2^{-(j+j'_Q)} - C2^j\right).$$

If $s_{jQ} \le k2^{-2j}$, we might as well put it to zero and obtain the bound $C(k^2 2^{-2j})^2$ for the integral. If $s_{jQ} > k2^{-2j}$, the integrand can be made homogeneous in $\mathbf{x}$ and the integral bounded by

$$\int_{r>s_{jQ}} \left(\frac{r}{k2^{-2j}}\right)^{-2Mp} r\,\mathrm{d}r \lesssim (s_{jQ})^2 \left(\frac{s_{jQ}}{k2^{-2j}}\right)^{-2Mp} \le \left(k2^{-2j}\right)^2 \left(\frac{s_{jQ}}{k2^{-2j}}\right)^{-2Mp+2}.$$

Now uniformly over s_{jQ}, the resulting bound is

$$C\left(k2^{-2j}\right)^2 \left(1 + 2^j \left(2^{-j'_Q} - k^{-1}2^{2j}\right)_+\right)^{-2Mp+2}. \tag{31}$$

The minimum of (30) and (31) is

$$C \min\left(2^{2j}, k^2 2^{-4j}\right) \left(1 + 2^j \left(2^{-j'_Q} - k^{-1}2^{2j}\right)_+\right)^{-2Mp+2}. \tag{32}$$

– If $k \le 2^{2j}$, then $\beta = 2^{-2j}$. This time we will only consider the lower bound for $\|\mathbf{x}\|_\infty$, and write

$$\int_{\|\mathbf{x}\|_\infty \ge C2^j} \left(1 + \left\|2^{-j}k\nabla\phi(x_Q) - \mathbf{x}\right\|^2\right)^{-Mp} \mathrm{d}\mathbf{x}.$$

Since $\|\nabla\phi(x_Q)\|$ is $O(1)$ and $2^{-j}k \le 2^j$, there exists a value $j^* \le \frac{1}{2}\log_2 k + C$ such that for all $j \ge j^*$, the center of the bump is inside the square $\|\mathbf{x}\|_\infty \le C2^j$ (the constant C changes from expression to expression). When this occurs, we can let $\mathbf{x}' = \mathbf{x} - 2^{-j}k\nabla\phi(x_Q)$ as before and consider the integral outside of a smaller square and bound

$$\int_{\|\mathbf{x}\|_\infty \ge C2^j} \left(1 + \|\mathbf{x}\|^2\right)^{-Mp} \mathrm{d}\mathbf{x} \le 2^{-2j(Mp-2)}, \quad j \ge j^*. \tag{33}$$

For the few values of j such that $\frac{1}{2}\log_2 k \le j \le j^*$, we recover the previous estimate, namely $C(k2^{-2j})^2$, which is $O(1)$.

- Consider now the sum over Q, and recall that the bounds just obtained need to be multiplied by $2^{-jp}(k\phi(x_Q))^{-p/2}$. Pull out the factor 2^{-jp} one more time. Again, we need to separately consider $2^{2j} \le k$ (large scales) and $2^{2j} \ge k$ (small scales). For small scales, the bound (33) is uniform in Q, hence the sum over $Q \in \mathcal{Q}_j$ simply contributes a factor 2^{2j}.

  For large scales, the strategy is to split the sum over Q into a *near-field contribution*, for which $d(s,t) \le C_1$ in the sense of Lemma 5, and a *far-field contribution*, for which stationary-phase points must be handled adequately. The terms in the far-field sum are then further broken down into groups corresponding to the same value of the scale defect, which Lemma 4 helps identify. Schematically,

$$\sum_{Q\in\mathcal{Q}_j} = \sum_{Q\in\text{near-field}} + \sum_{j'>0} \left[\sum_{Q:\text{scale defect}=j'}\right].$$

Consider the two regions separately.

– *Near-field*. In this region, $\phi(x_Q)$ may be as small as $1/k$, hence we estimate $(k\phi(x_Q))^{-p/2} \lesssim 1$. By Lemma 5, the scale defect j'_Q is bounded by a constant, hence the bound (32) becomes

$$C\min\bigl(2^{2j}, k^2 2^{-4j}\bigr)\bigl(1 + 2^j\bigl(1 - k^{-1}2^{2j}\bigr)_+\bigr)^{-2Mp+2}. \tag{34}$$

We claim that this quantity is always less than a constant independent of j and k. Indeed, if j is so large that $2^{2j}k^{-1} \geq 1/2$, then $k^2 2^{-4j} \leq 4$, and it suffices to use the trivial minoration $1 + 2^j(1 - k^{-1}2^{2j})_+ \geq 1$. If on the other hand $2^{2j}k^{-1} < 1/2$, then we have $1 + 2^j(1 - k^{-1}2^{2j})_+ \geq \frac{1}{2}2^j$, which implies that (34) is bounded by

$$C\min\bigl(2^{2j}, k^2 2^{-4j}\bigr)2^{-2j(Mp-2)} \leq C2^{-2j(-1+Mp-2)} \leq 1,$$

because we chose $Mp \geq 5$. The sum over Q then contributes a factor proportional to the number of non-diagonal squares, i.e., 2^{2j}.

– *Far-field*. The leading factor $(k\phi(x_Q))^{-p/2}$ now contributes a factor $k^{-p/2}$, since $\phi(x_Q) \geq C$ in the far-field. For the sum over Q, we use equation (32) one more time and write

$$C\min\bigl(2^{2j}, k^2 2^{-4j}\bigr) \sum_{Q\in\text{far-field}} \bigl(1 + 2^j\bigl(2^{-j'_Q} - k^{-1}2^{2j}\bigr)_+\bigr)^{-Mp+2}.$$

For each Q, find the closest integer $j' \leq j$ to j'_Q. As long as $j' < j$, Lemma 4 asserts that the number of terms comparable to $(1 + 2^j(2^{-j'} - k^{-1}2^{2j})_+)^{-Mp+2}$ is $O(2^{2j-j'})$. The endpoint $j' = j$ receives the contribution of arbitrary large j'_Q, meaning terms that can be as large as $O(1)$; however by Lemma 4 there can only be $O(2^j)$ such terms. After indexing terms by j' in place of Q, we get the bound

$$C\min\bigl(2^{2j}, k^2 2^{-4j}\bigr)\left[2^j + \sum_{-C\leq j'<j} 2^{2j-j'}\bigl(1 + 2^j\bigl(2^{-j'} - k^{-1}2^{2j}\bigr)_+\bigr)^{-Mp+2}\right].$$

It is easy to see that the summand peaks for j' near $j_0 = \min(j, -2j + \log_2 k)$; it decreases geometrically for $j \leq j_0$ because of the factor in brackets, and decreases geometrically for $j \geq j_0$ because of the factor $2^{2j-j'}$. The result is a bound

$$C\min\bigl(2^{2j}, k^2 2^{-4j}\bigr)\max\bigl(2^{4j}k^{-1}, 2^j\bigr) = C\min\bigl(k, 2^{3j}\bigr).$$

- What remains after gathering the various bounds is a constant $C_{M,p}$ times

$$\begin{aligned}
&\sum_{j\leq \frac{1}{2}\log_2 k + C} 2^{-jp}2^{2j} \quad \text{(near-field)}\\
&+ \sum_{j\leq \frac{1}{2}\log_2 k + C} k^{-p/2}2^{-jp}\min\bigl(k, 2^{3j}\bigr) \quad \text{(far-field)}\\
&+ \sum_{j> \frac{1}{2}\log_2 k + C} 2^{2j}2^{-jp}2^{-2j(Mp-2)} \quad \text{(small scales)}.
\end{aligned}$$

The near-field contribution sums up to $C_p k^{1-p/2}$ as soon as $p < 2$. The far-field contribution is bounded by

$$C_p k^{-p/2}\left[\sum_{j\le\frac{1}{3}\log_2 k} 2^{j(3-p)} + k\sum_{j>\frac{1}{3}\log_2 k} 2^{-jp}\right]$$
$$\le C_p k^{-p/2} k^{1-p/3} \le C_p k^{1-p/2}.$$

With the choice $Mp \ge 5$, the contribution of "small scales" is negligible in contrast to the first two terms. $O(k^{1-p/2})$ is the desired growth rate in k, compatible with (19). This concludes the part of the proof related to non-diagonal squares.

2.6 Diagonal Kernel Fragments: Decay of Individual Coefficients

It is now assumed that the dyadic square Q overlaps with the diagonal strip $S = \{(s,t) : d(s,t) \lesssim 1/k\}$. Because of the singularity of the kernel at $s = t$, the integrations by parts cannot proceed as before. Inside S, the smoothness of the kernel is governed by the case $kx \lesssim 1$ in Lemma 1.

Further complications arise, depending on the value of the scale j.

- At scales $j \le \log_2 k + C$, a square Q intersecting with S is not entirely contained in S; in fact, a large portion of it lies in the non-diagonal portion $d(s,t) \gtrsim 1/k$. We call this portion (two triangles) the *off-strip contribution*. There are $O(2^j)$ such triangles.
- At scales $j \ge \log_2 k + C$, some squares may be contained inside the strip S without intersecting the diagonal $s = t$. These squares make up the *regular on-strip contribution*; there are $O(2^{2j}k^{-1})$ such squares.
- The remaining $O(2^j)$ squares or portions thereof, overlapping with the diagonal $s = t$, make up the *singular on-strip contribution*.

In order to smoothly cut off the strip S from dyadic squares, introduce $\sigma = s - t$ (defined modulo 1 in $[0,1]$) and $\tau = s + t$. By symmetry of the problem, one can consider the triangle $\{(s,t) \in Q : s \ge t\}$ and still call it Q, without loss of generality. We can therefore focus on $\sigma > 0$. If we properly select the coset relative to the modulo operation, we can also take (σ, τ) to smoothly parametrize the triangle Q. With these choices, the diagonal strip is $S = \{(\sigma,\tau) : \sigma \lesssim 1/k\}$. Consider now a smooth indicator $\rho(k\sigma)$ where ρ is a C^∞ positive function obeying

$$\rho(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1; \\ 0 & \text{if } x \ge 2. \end{cases}$$

Multiplying the integrand in (41) by $\rho(k\sigma)$ gives the on-strip contribution (regular and singular); multiplying it by $1 - \rho(k\sigma)$ gives the off-strip contribution. These cases are treated separately.

- *Off-strip contribution.* Over the off-strip region we have $k\sigma \gtrsim 1$ hence $k\phi(s,t) \gtrsim 1$ by (9), so that the case $kx \gtrsim 1$ of Lemma 1 applies there. The analysis of the coefficient decay in $\mathbf{n}$ is the same as in the previous section, so we omit it here.

As far as analysis of the decay in $\mathbf{m}$, we are back in the setting of the analysis of Sect. 2.4, except for the factor $1-\rho(k\sigma)$ that prevents the same scheme of integrations by parts in σ. Each derivative of $\rho(k\sigma)$ would produce an unacceptably large factor k. The smoothness in τ is however unaffected, which permits to carry over the analysis of Sect. 2.4 with integrations by parts in τ only. The direction of increasing τ in the $\mathbf{m}$ plane is $\mathbf{e}_\tau = (1,1)$, to which corresponds the decomposition $2\mathbf{m}\cdot(s,t) = (m_1+m_2)\tau + (m_1-m_2)\sigma$. Since $\nabla\phi(0) \equiv \lim_{\sigma\to 0^+}\nabla\phi$ points in the direction of σ, we have $\mathbf{e}_\tau\cdot\nabla\phi(0) = 0$. Repeated integrations by parts should now be carried out, in the fashion of Sect. 2.4, but with a different differential operator than L:

$$L_0 = \frac{I - \beta\frac{\partial^2}{\partial\tau^2} - \mathrm{i}\beta k(\frac{\partial^2\phi}{\partial\tau^2})}{1+\beta 2^{2j}(\frac{m_1+m_2}{2})^2},$$

where β is the same as previously. Notice that the amplitude is uniformly bounded, since $kx \gtrsim 1$ in the off-strip region. The result is a bound that involves m_1+m_2 only. With the contribution of the decay in $\mathbf{n}$, the off-strip coefficient estimate is

$$\left|\left\langle K_Q, \big(1-\rho(k\sigma)\big)\varphi_{j,\mathbf{m},\mathbf{n}}\right\rangle\right| \le C_M 2^{-j}\big(1+\beta 2^{2j}(m_1+m_2)^2\big)^{-M}\big(1+\|\mathbf{n}\|\big)^{-M}, \tag{35}$$

for all $M>0$, and only for scales obeying $2^{2j}\lesssim k$.

- *On-strip contribution: amplitude estimate.* We now take $k\sigma\lesssim 1$. The case $kx\lesssim 1$ of Lemma 1 allows to write

$$K(s,t) = \mathrm{e}^{\mathrm{i}k\phi(s,t)}a\big(k\phi(s,t),s,t\big),$$

where now the amplitude's smoothness is

$$\left|\frac{\mathrm{d}^n a}{\mathrm{d}\phi^n}(k\phi,s,t)\right| \le C_n\phi^{-n}, \quad \phi\lesssim 1. \tag{36}$$

The partial derivatives of a with respect to the arguments s and t are $O(1)$ and well within the above bound as long as $\phi(s,t)\lesssim 1$. To compute the *total* derivatives with respect to s and t, however, it is necessary to contrast smoothness along and across the oscillations, by means of the coordinates σ and τ. The value of $\phi(s,t)$ is comparable to the circle distance $d(\sigma,0)$, namely

$$Dd(\sigma,0) \le \phi\left(\frac{\sigma+\tau}{2},\frac{\tau-\sigma}{2}\right) \le \tilde{D}d(\sigma,0),$$

for some fixed D and $\tilde{D}$. The first inequality is exactly (9), the last inequality follows from a Taylor expansion. Since we only consider $\sigma>0$, we write this property as $\phi\asymp\sigma$. A careful analysis of Taylor remainders shows that the same estimate is true for the τ derivatives,

$$\left|\frac{\mathrm{d}^n\phi}{\mathrm{d}\tau^n}\left(\frac{\sigma+\tau}{2},\frac{\tau-\sigma}{2}\right)\right| \le C_n\sigma, \quad \sigma\neq 0, \tag{37}$$

while the σ derivatives do not yield any gain:

$$\left|\frac{\mathrm{d}^n\phi}{\mathrm{d}\sigma^n}\left(\frac{\sigma+\tau}{2},\frac{\tau-\sigma}{2}\right)\right|\le C_n,\quad \sigma\neq 0. \tag{38}$$

The action of the successive τ derivatives on a through its ϕ dependence can be understood from the higher-order analog of the chain rule, known as the combinatorial Faà di Bruno formula:

$$\left(\frac{\mathrm{d}}{\mathrm{d}\tau}\right)^n a\left(k\phi\left(\frac{\sigma+\tau}{2},\frac{\tau-\sigma}{2}\right),s_0,t_0\right)=\sum_{\pi\in\Pi}\left(\frac{\mathrm{d}^{|\pi|}a}{\mathrm{d}\phi^{|\pi|}}\right)\cdot\prod_{B\in\pi}k\frac{\mathrm{d}^{|B|}\phi}{\mathrm{d}\tau^{|B|}}.$$

In this formula, Π is the set of all partitions π of $\{1,\dots,n\}$; $|\pi|$ denotes the number of blocks in the partition π; these blocks are indexed as $B\in\pi$; and $|B|$ denotes the size of the block B. Since there are $|\pi|$ factors in the product over B, (37) reveals that the derivatives of ϕ yield a factor $C_n\sigma^{|\pi|}$. On the other hand, by (36), each ϕ-differentiation of a introduces an inverse power of σ. The order of the derivative is $|\pi|$, for a contribution of $\sigma^{-|\pi|}$ that exactly cancels the $\sigma^{|\pi|}$ coming from the derivatives of ϕ.

This analysis only concerns the dependence of a on τ via ϕ. It is easy to apply the multivariable chain rule to see that the dependence of a on τ via its second and third arguments (s and t) does not change the conclusion that any number n of τ derivatives, $n\ge 1$, keep the amplitude bounded, with bound independent of j and k (but not n, of course):

$$\left|\frac{\mathrm{d}^n a}{\mathrm{d}\tau^n}\left(k\phi\left(\frac{\sigma+\tau}{2},\frac{\tau-\sigma}{2}\right),\frac{\sigma+\tau}{2},\frac{\tau-\sigma}{2}\right)\right|\le C_n,\quad n\ge 1,\sigma\neq 0. \tag{39}$$

There are no factors to gain in the σ derivatives of the phase, hence the same analysis yields

$$\left|\frac{\mathrm{d}^n a}{\mathrm{d}\sigma^n}\left(k\phi\left(\frac{\sigma+\tau}{2},\frac{\tau-\sigma}{2}\right),\frac{\sigma+\tau}{2},\frac{\tau-\sigma}{2}\right)\right|\le C_n\phi^{-n},\quad \phi\lesssim 1,n\ge 1,\sigma\neq 0. \tag{40}$$

We are now equipped to study the coefficient

$$\begin{aligned}\langle K_Q,\varphi_{j,\mathbf{m},\mathbf{n}}\rangle=\int_{3Q} w\big(2^j s,2^j t\big)\mathrm{e}^{\mathrm{i}k\phi(s,t)}a\big(k\phi(s,t),s,t\big)\mathrm{e}^{-\mathrm{i}2^j\mathbf{m}\cdot(s,t)}\\ \times 2^j\varphi_{(j,\mathbf{m})}\big(2^j s-n_1,2^j t-n_2\big)\,\mathrm{d}s\,\mathrm{d}t.\end{aligned} \tag{41}$$

- *Regular on-strip contribution*

  For regular on-strip squares, i.e., those squares at very small scales $j\ge\log_2 k+C$ that intersect with the strip S but not with the diagonal $s=t$, the decay in $\mathbf{m}$ and $\mathbf{n}$ is obtained by a simple argument of integration by parts. In contrast to the operators L and L_0 used earlier, we should now introduce copies of

$$L_1=\frac{I-2^{-2j}\Delta_{(\sigma,\tau)}}{1+\|\mathbf{m}\|^2},$$

and integrate by parts in (41). Each derivative in σ acting on the amplitude $a(k\phi, s, t)$ produces a factor $\phi^{-1} \asymp \sigma^{-1}$. Since Q does not intersect with the diagonal, $\sigma \gtrsim 2^{-j}$ hence $\sigma^{-1} \lesssim 2^{j}$. This factor is balanced by the choice of scaling in the expression of L_2. A fortiori, the derivatives in τ are governed by a stronger estimate and are therefore under control. The derivative in τ or σ acting on $\varphi_{(j,\mathbf{m},\mathbf{n})}$ do not compromise its super-algebraic decay, hence we gather the same decay in $\mathbf{n}$ as previously. One complication is however the possible logarithmic growth near $\sigma = 0$ of the amplitude a when it is not differentiated. Consider the intermediate bound

$$\begin{aligned}&\big|\langle K_Q, \rho(k\sigma)\varphi_{j,\mathbf{m},\mathbf{n}}\rangle\big| \\ &\quad\le C_M 2^j \big(1+\|\mathbf{m}\|^2\big)^{-M}\big(1+\|\mathbf{n}\|\big)^{-M}\int_{3Q}\big|a\big(k\phi(s,t),s,t\big)\big|\,\mathrm{d}s\,\mathrm{d}t,\end{aligned}$$

for all $M > 0$. If the amplitude were bounded, then the integral would produce a factor 2^{-2j} as in the non-diagonal case. Instead, we claim that the integral factor is bounded by $2^{-j}k^{-1}$. In order to see this, consider the bound

$$\big|a(k\phi, s, t)\big| \le C\big(1+\big|\log\big(k\phi(s,t)\big)\big|\big),$$

from Lemma 1. Since log is increasing and $\phi \asymp \sigma$, there exist $C_1, C_2 > 0$ such that

$$\log(C_1 k\sigma) \le \log\big(k\phi(s,t)\big) \le \log(C_2 k\sigma),$$

hence

$$\big|\log\big(k\phi(s,t)\big)\big| \le C + \big|\log(k\sigma)\big|, \quad \text{for some } C > 0.$$

This bound does not depend on τ, and since $(s,t) \in 3Q$, τ ranges over a set of length $O(2^{-j})$. We therefore obtain the bound

$$\int_{3Q}\big|a\big(k\phi(s,t),s,t\big)\big|\,\mathrm{d}s\,\mathrm{d}t \le C2^{-j} \times \int_0^{\frac{1}{Ck}}\big(C+\big|\log(k\sigma)\big|\big)\,\mathrm{d}\sigma \le C2^{-j}k^{-1}.$$

The final estimate for the regular on-strip contribution is

$$\big|\langle K_Q, \rho(k\sigma)\varphi_{j,\mathbf{m},\mathbf{n}}\rangle\big| \le C_M k^{-1}\big(1+\|\mathbf{m}\|^2\big)^{-M}\big(1+\|\mathbf{n}\|\big)^{-M} \quad \text{(regular on-strip)}. \tag{42}$$

- *Singular on-strip contribution*

  Let us now consider a dyadic square Q that intersects with the diagonal $s = t$. In the language of microlocal analysis, $s = t$ is the singular support of the kernel, and we expect the corresponding wavefront set $\{(s, s, \xi, -\xi)\}$ to play a role in the analysis. Accordingly, we show in this section that large wave atom coefficients cluster around the wavefront set. In particular, we cannot expect that the decay length scale of the wave atom coefficients be independent of j in all directions in $\mathbf{m}$: because Q overlaps with, or is close to the diagonal, the decay in the direction $m_1 - m_2$ (perpendicular to the diagonal) is much slower than the decay in the direction $m_1 + m_2$ (parallel to the diagonal). However, the number of diagonal squares

is small enough to restore the overall balance at the level of the ℓ_p summability criterion.

To quantify the decay in the $m_1 + m_2$ direction, introduce the self-adjoint operator

$$L_2 = \frac{I - 2^{-2j}\frac{\partial^2}{\partial \tau^2}}{1 + (\frac{m_1+m_2}{2})^2}.$$

It leaves the exponential $\mathrm{e}^{-\mathrm{i}2^j(m_1 s + m_2 t)}$ invariant. After integrating by parts, the action of $I - 2^{-2j}\frac{\partial^2}{\partial \tau^2}$ leaves the bound on the rest of the integrand unchanged, because

(i) $w(2^j s, 2^j t)$ and $\varphi_{(j,\mathbf{m})}(2^j s - n_1, 2^j t - n_2)$ produce a factor 2^j when differentiated.
(ii) Differentiating $\varphi_{(j,\mathbf{m})}(2^j s - n_1, 2^j t - n_2)$ does not compromise its decay in $\mathbf{n}$.
(iii) $a(k\phi, s, t)$ has a logarithmic singularity, and otherwise becomes uniformly bounded when differentiated in τ, as we have seen. (The presence of the scaling 2^{-2j} in L_1 is not even needed here.)

The integral $\int_{3Q} |a(k\phi(s,t), s, t)|\,\mathrm{d}s\,\mathrm{d}t$ for the amplitude can be bounded by $C2^{-j}\max(2^{-j}, k^{-1})$ as we argued for the regular on-strip squares (here $3Q$ is not necessarily contained in S). The result is a bound

$$\begin{aligned}&\left|\left\langle K_Q, \rho(k\sigma)\varphi_{j,\mathbf{m},\mathbf{n}}\right\rangle\right| \\ &\quad \le C_M \max\left(2^{-j}, k^{-1}\right)\left(1 + (m_1 + m_2)^2\right)^{-M}\left(1 + \|\mathbf{n}\|\right)^{-M} \\ &\quad \text{(singular on-strip)}\end{aligned} \tag{43}$$

for all $M > 0$.

Finally, the decay in $m_1 - m_2$ for those (singular, on-strip) squares that intersect the diagonal cannot proceed as previously. An analysis of coefficients taken individually would be far from sharp, e.g., would not even reproduce ℓ_2 summability. The proper reasoning involves a *collective bound* on the ℓ_2 norm of all the wave atom coefficients at a given scale $j > 0$, which correspond to squares Q that intersect with the diagonal. This reasoning is explained in the next section, and gives the bound

$$\begin{aligned}&\sum_{\mathbf{m},\mathbf{n}} \left|\left\langle K_Q, \rho(k\sigma)\varphi_{j,\mathbf{m},\mathbf{n}}\right\rangle\right|^2 \le Cj^2 2^{-3j}, \\ &\qquad Q \in \mathcal{Q}_j \text{ and } Q \text{ intersects the diagonal.}\end{aligned} \tag{44}$$

The study of ℓ_p summability from all these estimates is then treated in Sect. 2.8.

2.7 Diagonal Kernel Fragments: Collective Decay Properties

The strategy for obtaining (44) is to compare wave atom coefficients to wavelet coefficients, scale by scale. Estimating individual wavelet coefficients is a much tighter

way to capture the sparsity of a log singularity than directly through wave atoms. (Wavelets however are not well-adapted for the overwhelming majority of dyadic squares that correspond to C^∞ oscillations.)

Consider two-dimensional compactly supported Daubechies wavelets with one dilation index j, built on the principle of multiresolution analysis [19]. They are denoted as

$$\psi^\varepsilon_{j',\mathbf{n}}(s,t) = 2^{j'}\psi^\varepsilon\big(2^{j'}s - n_1, 2^{j'}t - n_2\big),$$

where $\varepsilon = 1, 2, 3$ indexes the type of the wavelet (HH, HL or LL). The easiest way to define Meyer wavelets in a square is to periodize them at the edges.

Fix $j \geq 0$, consider a function $f(s,t)$ defined in $[0,1]^2$, and consider its wave atom coefficients at a scale j. By Plancherel for wave atoms, there exists an annulus $A_j = \{(\xi_1,\xi_2) : C_1 2^{2j} \leq \|\xi\|_\infty \leq C_2 2^{2j}\}$ such that

$$\sum_{\mathbf{m},\mathbf{n}} \big|\langle f, \varphi_{j,\mathbf{m},\mathbf{n}}\rangle\big|^2 \leq \int_{A_j} \big|\hat f(\xi)\big|^2 \,\mathrm{d}\xi.$$

By Plancherel for wavelets and the properties of Daubechies wavelets [19], there exists j_0 such that this L^2 energy is for the most part accounted for by the wavelet coefficients at scales $2j - j_0 \leq j' \leq 2j + j_0$, i.e.,

$$\frac{1}{2}\int_{A_j} \big|\hat f(\xi)\big|^2 \,\mathrm{d}\xi \leq \sum_{j'\in[2j-j_0,2j+j_0]} \sum_{\varepsilon,\mathbf{n}} \big|\big\langle f, \psi^\varepsilon_{j',\mathbf{n}}\big\rangle\big|^2. \tag{45}$$

The last two equations show that, collectively in an ℓ_2 sense, wave atom coefficients at scale j can be controlled by wavelet coefficients at scales neighboring $2j$.

The relevant range of scales for this analysis is $j \geq \frac{1}{2}\log_2 k$. The on-strip region has length $O(2^{-j})$ and width $O(\min(k^{-1}, 2^{-j}))$. Since each wavelet is supported in a square of size $\sim 2^{-2j}$-by-2^{-2j}, the number of wavelets that intersect the strip is $O(2^j \times 2^{2j}\min(k^{-1}, 2^{-j})) = O(\min(2^{3j}k^{-1}, 2^j))$. Among those, only $O(2^j)$ correspond to wavelets intersecting with the diagonal $s = t$. The bound on wavelet coefficient depends on their location with respect to the diagonal:

- *Non-diagonal wavelets*. The wavelet's wave number is $\sim 2^{j'} \sim 2^{2j}$ and soon becomes much larger than the local wave number $\sim k$ of the oscillations of the kernel, hence a fast decay in $j' \to \infty$. More precisely, fix $Q \in \mathcal{Q}_j$; the coefficient of interest is

$$\big\langle K_Q \rho(k\sigma), \psi^\varepsilon_{j',\mathbf{n}}\big\rangle = \int_{\mathrm{supp}\,\psi^\varepsilon_{j',\mathbf{n}}} a\big(k\phi(s,t),s,t\big)\mathrm{e}^{\mathrm{i}k\phi(s,t)} w\big(2^j s, 2^j t\big)2^{j'} \times \psi^\varepsilon\big(2^{j'}s - n_1, 2^{j'}t - n_2\big)\,\mathrm{d}s\,\mathrm{d}t.$$

  Because the wavelet has at least one vanishing moment, one may write it either as

$$\psi^\varepsilon\big(2^{j'}s - n_1, 2^{j'}t - n_2\big) = 2^{-j'}\frac{\mathrm{d}\tilde\psi^\varepsilon}{\mathrm{d}s}\big(2^{j'}s - n_1, 2^{j'}t - n_2\big), \quad \varepsilon = \text{HL or HH},$$

or as

$$\psi^{\varepsilon}\bigl(2^{j'}s-n_1,2^{j'}t-n_2\bigr)=2^{-j'}\frac{\mathrm{d}\tilde{\psi}^{\varepsilon}}{\mathrm{d}t}\bigl(2^{j'}s-n_1,2^{j'}t-n_2\bigr),\quad \varepsilon=\mathrm{LH},$$

where $\tilde{\psi}^{\varepsilon}$ has the same support as ψ^{ε}. After integrating by parts in s or t, we can
– use the bounds (39) and (40) on the amplitude
– use the bound $\nabla_{(s,t)}\mathrm{e}^{\mathrm{i}k\phi}=O(k)$
– use $\nabla_{(s,t)}w(2^j s,2^j t)=O(2^j)$
– use $|\operatorname{supp}\psi^{\varepsilon}_{j',\mathbf{n}}|\lesssim 2^{-2j'}$
to conclude that the coefficient obeys

$$\bigl|\bigl\langle K_Q\rho(k\sigma),\psi^{\varepsilon}_{j',\mathbf{n}}\bigr\rangle\bigr|\le C_M 2^{-j'}2^{-j'}\bigl(\max\bigl(k,2^j\bigr)+\overline{\phi^{-1}}\bigr),$$

where $\overline{\phi^{-1}}$ is a notation for the supremum of ϕ^{-1} over the support of the wavelet. If we index by the integer $q\ge 1$ the distance between the center of the support of the wavelet to the diagonal, as $\sqrt{2}q2^{-j'}$, then $\overline{\phi^{-1}}\asymp q^{-1}2^{j'}$. The bound above becomes $C_M 2^{-j'}(q^{-1}+2^{-j'}\max(k,2^j))$.

The sum in the right-hand side of (45) is then estimated as follows. As we saw earlier the length of the strip $S\cap Q$ is $O(2^{-j})$, and its width is $\min(1/k,2^{-j})$. Since the translation step of wavelets is $2^{-j'}\sim 2^{-2j}$, the translation index $\mathbf{n}$ takes on $2^j\times 2^{2j}\min(1/k,2^{-j})$ values. Hence

$$\begin{aligned}\sum_{j'\in[2j-j_0,2j+j_0]}\sum_{\varepsilon,\mathbf{n}}\bigl|\bigl\langle K_Q\rho(k\sigma),\psi^{\varepsilon}_{j',\mathbf{n}}\bigr\rangle\bigr|^2 &\le C\sum_{\mathbf{n}}\bigl|2^{-2j}\bigl(q^{-1}+2^{-2j}k\bigr)\bigr|^2\\ &\le C2^j\sum_{q\ge 1}\bigl(2^{-4j}q^{-2}\bigr)+2^{3j}\min\biggl(\frac{1}{k},2^{-j}\biggr)\\ &\quad\times 2^{-4j}\bigl(2^{-2j}\max\bigl(k,2^j\bigr)\bigr)^2\\ &\le C2^{-3j}\quad\biggl(\text{because } j\ge\frac{1}{2}\log_2 k+C\biggr).\end{aligned}$$

- *Diagonal wavelets*. For wavelets intersecting the diagonal, it will not be necessary to quantify cancellations. By Lemma 1,

$$\begin{aligned}&\bigl|\bigl\langle K_Q\rho(k\sigma),\psi^{\varepsilon}_{j',\mathbf{n}}\bigr\rangle\bigr|\\ &\quad\le C\int_{\operatorname{supp}\psi^{\varepsilon}_{j',\mathbf{n}}}\bigl(1+\bigl|\log(k\sigma)\bigr|\bigr)2^{j'}\bigl|\psi^{\varepsilon}\bigl(2^{j'}s-n_1,2^{j'}t-n_2\bigr)\bigr|\,\mathrm{d}s\,\mathrm{d}t.\end{aligned}\tag{46}$$

Without loss of generality we can consider $k\sigma<1/2$ and write

$$\bigl|\log_2(k\sigma)\bigr|=-\log_2(k\sigma)=-\log_2\bigl(2^{j'}\sigma\bigr)-\log_2 k+j'\le-\log_2\bigl(2^{j'}\sigma\bigr)+j'.$$

Since log is integrable near the origin, and $|\operatorname{supp}\psi^{\varepsilon}_{j',\mathbf{n}}|\lesssim 2^{-2j'}$, the contribution due to $-\log_2(2^{j'}\sigma)$ is $O(2^{-j'})$. The contribution of the lone j', on the other hand, is $O(j'2^{-j'})$.

There are $O(2^j)$ diagonal wavelets, each with coefficients $O(j'2^{-j'}) = O(j2^{-2j})$, hence the sum of their squares in the range $2j - j_0 \le j' \le 2j + j_0$ is $O(2^j \times (j2^{-2j})^2) = O(j^2 2^{-3j})$.

As $j \to \infty$, the contribution of diagonal wavelets manifestly dominates that of non-diagonal wavelets, and we have shown that the resulting estimate is (44).

2.8 Diagonal Kernel Fragments: ℓ_p Summation

Let us conclude by calculating the growth of $\sum_j \sum_{Q\in\mathcal{Q}_j} \sum_{\mathbf{m},\mathbf{n}} |\langle K_Q, \varphi_{j,\mathbf{m},\mathbf{n}}\rangle|^p$ in the parameter k, for those dyadic squares that intersect the strip $k\sigma \lesssim 1$. We start by letting $p \le 1$.

Consider each contribution separately.

- *Off-strip contribution*. Recall that $j \le \frac{1}{2}\log_2 k + C$ in this case. The reasoning entirely parallels that of the previous section and we encourage the reader to focus on the discrepancies. First, the sum over $\mathbf{n}$ yields a harmless constant factor. Second, use (35) and drag the factor 2^{-jp} out of the sums over $\mathbf{m}$ and Q; the former sum is then comparable to the integral

$$I = \int_{C_j} \left(1 + \left|\beta^{1/2} 2^j (x_1 + x_2)\right|^2\right)^{-Mp} \mathrm{d}x_1\, \mathrm{d}x_2,$$

  with C_j an annulus of inner and outer radii proportional to 2^j. Over this domain, the integrand concentrates near the union of two "ridges" of length $\sim 2^j$ and width $\sim \beta^{-1/2} 2^{-j}$, oriented along the anti-diagonal $x_1 = -x_2$. Note that $\beta = 2^{2j}k^{-2}$. The integral I is therefore bounded by a constant times $2^j \times (\beta^{-1/2}2^{-j}) = 2^{-j}k$. Third, the sum over $Q \in \mathbf{Q}_j$ that intersect with the strip yields a factor 2^j, proportional to the number of diagonal dyadic squares at scale j. The remaining sum is bounded by

$$C_p \sum_{j \le \frac{1}{2}\log_2 k + C} 2^j \left(2^{-j}k\right) 2^{-jp} \le C_p k^{1-p/2}.$$

  This is the desired growth rate in k.
- *Regular on-strip contribution*. Here, $j \ge \frac{1}{2}\log_2 k + C$, therefore $\|\mathbf{m}\|_\infty \ge C2^j \ge C\sqrt{k}$. The factor $(1 + \|\mathbf{m}\|_\infty)^{-M}$ in (42) therefore yields a negative power $k^{-M/2}$ for all $M > 0$, i.e., what we denoted earlier as $k^{-\infty}$. This is negligible in comparison with $k^{1-p/2}$.
- *Singular on-strip contribution*. As previously, two scale regimes should be considered. When $j \le \frac{1}{2}\log_2 k$, we can use the bound (43). The sum over $\mathbf{n}$ is harmless; the sum over $\mathbf{m}$ produces a factor 2^j since there are significant $O(2^j)$ values of $\mathbf{m}$ on the ridge $|m_1 + m_2| \le C$ at scale j; the sum over Q produces another factor 2^j since there are $O(2^j)$ diagonal dyadic squares at scale j. The resulting sum over j is then bounded by

$$C_p \sum_{j \le \frac{1}{2}\log_2 k} 2^{2j} 2^{-jp} \le C_p k^{1-p/2},$$

  which is again the desired growth rate.

If now $j \geq \frac{1}{2}\log_2 k$, we need to invoke the collective decay estimate (44). Fix j and $Q \in \mathcal{Q}_j$ a singular dyadic square. By (43), values of $m_1 + m_2$ significantly different from zero will give rise to negligible coefficients that sum up to $o(k)$. More precisely, let $\delta > 0$ be arbitrarily small. Then the wave atom coefficients in the region $|m_1 + m_2| \geq C2^{\delta j}$ decay sufficiently fast (take $M \gg 1/\delta$) that their total contribution is $O(k^{-\infty})$ in ℓ_p. The significant coefficients at scale j are, again, on a ridge of length $O(2^j)$ and width $O(2^{\delta j})$, for a combined total of $N = O(2^{j(1+\delta)})$ significant coefficients.

We can now relate the ℓ_2 norm estimate (44) to an ℓ_p estimate, $0 < p < 2$, by means of the Hölder inequality

$$\left(\sum_{\mathbf{m},\mathbf{n}} \left|\left\langle K_Q \rho(k\sigma), \varphi_{j,\mathbf{m},\mathbf{n}}\right\rangle\right|^p\right)^{1/p} \leq \left(\sum_{\mathbf{m},\mathbf{n}} \left|\left\langle K_Q \rho(k\sigma), \varphi_{j,\mathbf{m},\mathbf{n}}\right\rangle\right|^2\right)^{1/2} \times N^{\frac{1}{p}-\frac{1}{2}}, \tag{47}$$

where $N = O(2^{j(1+\delta)})$. After simplification, the right-hand side is bounded by $C_{\delta,p} j 2^{j(-2+1/p+\delta')}$ where δ' is another arbitrarily small number, namely $\delta' = \delta(\frac{1}{p} - \frac{1}{2})$. This quantity still needs to be summed over Q—there are $O(2^j)$ such squares—and then over j; but the summation method will depend how p compares to 1, and accordingly, which of (24) or (25) should be used.

If $p \leq 1$, then (24) should be used, and we obtain

$$\|K_\mu\|_{\ell_p(F)}^p \leq C_{\delta,p} \sum_{j \geq \frac{1}{2}\log_2 k} 2^j \left(j 2^{j(-2+\frac{1}{p}+\delta')}\right)^p = C_{\delta,p} \sum_{j \geq \frac{1}{2}\log_2 k} j^p 2^{j(2-2p+\delta' p)},$$

which always diverges. However if $p \geq 1$, then (25) implies

$$\|K_\mu\|_{\ell_p(F)} \leq C_{\delta,p} \sum_{j \geq \frac{1}{2}\log_2 k} 2^j \left(j 2^{j(-2+\frac{1}{p}+\delta')}\right).$$

For any $1 < p < 2$, the above series is convergent if for instance we choose $\delta' = \frac{1}{2}(1 - \frac{1}{p})$, and the result is $O(1)$, independent of k. This part of the singular on-strip contribution falls into the second category identified at the beginning of the proof, namely (20). This concludes the proof in the case when K is the single-layer potential G_0.

2.9 Analysis of the Double-Layer Potential

The proof of the sparsity result for the double-layer kernel G_1 defined in (11) is a simple modification of that for the single-layer kernel G_0.

- *Non-diagonal part.* The smoothness bound for Hankel functions in Lemma 1 exhibits the same rate for all $n \geq 0$ in the case when $x \gtrsim 1/k$. The other factors, functions of s and t which accompany the Hankel factor in formula (11), have no bearing on the sparsity analysis since they are smooth and do not depend on k but for the leading factor $ik/4$. As a consequence, G_1 can still be written as a product

$$G_1(s,t) = a\big(k\phi(s,t), s, t\big) \mathrm{e}^{\mathrm{i}k\phi(s,t)} \quad \big(\text{non-diagonal part}, |s-t| \gtrsim 1/k\big),$$

where the amplitude obeys the same estimate as previously, but for a factor k:

$$\left|\frac{\mathrm{d}^n}{\mathrm{d}\phi^n}a\big(k\phi(s,t),s,t\big)\right| \le C_n k\frac{1}{\sqrt{k\phi(s,t)}}\phi(s,t)^{-n}.$$

With a number of non-standard wave atom coefficients $|\Lambda| = O(k\epsilon^{-1/\infty})$, one could form an approximation of the non-diagonal part of G_0 with error ϵ. In the case of G_1, the same number of terms results in an error that we can only bound by $k\epsilon$. Thus, to make the error less than a specified $\tilde{\epsilon}$, the number of terms needs to be $O(k^{1+1/\infty}\tilde{\epsilon}^{-1/\infty})$. This justifies the form of the first term in (14).

- *Diagonal part.* For $x \lesssim 1/k$ the smoothness estimate in Lemma 1 is worse for $H_1^{(1)}$ than for $H_0^{(1)}$, but as is well-known, the dot product

$$\frac{\mathbf{x}(s)-\mathbf{x}(t)}{\|\mathbf{x}(s)-\mathbf{x}(t)\|}\cdot n_{\mathbf{x}(t)}\|\dot{\mathbf{x}}(t)\|$$

is small near the diagonal and more than compensates for the growth of $H_1^{(1)}$ there. The precise version of this heuristic is a decomposition

$$G_1(s,t) = a\big(k\phi(s,t),s,t\big)\mathrm{e}^{\mathrm{i}k\phi(s,t)} \quad \big(\text{diagonal strip, } |s-t| \lesssim 1/k\big),$$

where we claim that the amplitude obeys the same estimates as those for G_0 in variables $\sigma = s-t$, $\tau = s+t$, namely

$$\left|\frac{\mathrm{d}^n a}{\mathrm{d}\tau^n}\left(k\phi\left(\frac{\sigma+\tau}{2},\frac{\tau-\sigma}{2}\right),\frac{\sigma+\tau}{2},\frac{\tau-\sigma}{2}\right)\right| \le C_n, \tag{48}$$

$$\left|\frac{\mathrm{d}^n a}{\mathrm{d}\sigma^n}\left(k\phi\left(\frac{\sigma+\tau}{2},\frac{\tau-\sigma}{2}\right),\frac{\sigma+\tau}{2},\frac{\tau-\sigma}{2}\right)\right| \le C_n\phi^{-n}, \quad \phi \lesssim 1, \tag{49}$$

for $\sigma \lesssim 1/k$, and this time for every $n \ge 0$ including zero. (Total derivatives in σ are defined by keeping τ fixed, and vice-versa.)

Let us prove (48). As previously, put $r = \frac{\mathbf{x}(s)-\mathbf{x}(t)}{\|\mathbf{x}(s)-\mathbf{x}(t)\|}$. Observe that $n_{\mathbf{x}(t)}\|\dot{\mathbf{x}}(t)\| = (\dot{\mathbf{x}}(t))^{\perp}\cdot r$. Derivatives of $(\dot{\mathbf{x}}(t))^{\perp}\cdot r$ in σ and τ are treated by the following lemma, proved in the Appendix.

Lemma 6 *For all $n \ge 0$,*

$$\left|\frac{\mathrm{d}^n}{\mathrm{d}\tau^n}\big[\big(\dot{\mathbf{x}}(t)\big)^{\perp}\cdot r\big]\right| \le C_n\sigma, \tag{50}$$

$$\left|\frac{\mathrm{d}^n}{\mathrm{d}\sigma^n}\big[\big(\dot{\mathbf{x}}(t)\big)^{\perp}\cdot r\big]\right| \le C_n\sigma^{1-n}. \tag{51}$$

The chain rule and Faà di Bruno formula can then be invoked as previously, with the combined knowledge of (50), (51), the growth of $H_1^{(1)}$ from Lemma 1, i.e.,

$$\left|\frac{\mathrm{d}^m}{\mathrm{d}\phi^m}\big[H_1^{(1)}\big(k\phi(s,t)\big)\mathrm{e}^{-\mathrm{i}k\phi(s,t)}\big]\right| \le C_m\frac{1}{k\sigma}\sigma^{-m},$$

as well as (37) and (38) on the growth of the derivatives of ϕ. It is straightforward to see that (48) is satisfied; for instance, the factor σ from $(\dot{\mathbf{x}}(t))^{\perp}\cdot r$ and the leading k in the expression of G_1 cancel out the $1/(k\sigma)$ in the formula for the derivatives of $H_1^{(1)}$. The rest of the argument involving the Faà di Bruno formula is the same as previously. Equation (49) follows from the same reasoning, and the observation that no σ factor is gained upon differentiating ϕ in σ.

Since (48) and (49) are at least as good as what they were in the case of G_0, the rest of the argument can proceed as previously with the same results; the "off-strip" and "regular on-strip" contributions, for instance, are unchanged from the G_0 scenario. The "singular on-strip" contribution however, corresponding to dyadic squares that intersect the diagonal, ought to be revisited since G_1 has a much milder singularity than G_0 near the diagonal.

The estimate of fast decay in $|m_1+m_2|$ and $\|\mathbf{n}\|$, namely (43), is a fortiori still valid. It appears, however, that the collective bound (44) at scale j can be improved to

$$\sum_{\mathbf{m},\mathbf{n}} \left|\left\langle K\chi_j^{\mathrm{diag}}\rho(k\sigma),\varphi_{j,\mathbf{m},\mathbf{n}}\right\rangle\right|^2 \le C2^{-6j}k^2, \tag{52}$$

where $\chi_j^{\mathrm{diag}}(s,t)$ refers to the $\sum_Q w_Q(s,t)$ over the squares $Q\in\mathcal{Q}_j$ at scale j for which the support of w_Q intersects the diagonal. The presence of an aggregation of windows $\chi_j^{\mathrm{diag}}(s,t)$ is important here, as the study of coefficients corresponding to individual windows w_Q would not give a sharp bound. Whether $\chi_j^{\mathrm{diag}}(s,t)$ or $\rho(k\sigma)$ effectively determines the cutoff depends on the relative values of j and $\log_2 k$.

Again, via a Plancherel argument, the scale-by-scale bound (52) can be proved by passing to a system of Daubechies wavelets. We have

$$\begin{aligned}\left\langle K\chi_j^{\mathrm{diag}}\rho(k\sigma),\psi^{\varepsilon}_{j',\mathbf{n}}\right\rangle = \int_{\mathrm{supp}\,\psi^{\varepsilon}_{j',\mathbf{n}}} &\chi_j^{\mathrm{diag}}(s,t)\rho(k\sigma)\frac{\mathrm{i}k}{4}H_1^{(1)}\big(k\phi(s,t)\big)\big(\dot{\mathbf{x}}(t)\big)^{\perp}\cdot r\\ &\times 2^{j'}\psi^{\varepsilon}\big(2^{j'}s-n_1,2^{j'}t-n_2\big)\,\mathrm{d}s\,\mathrm{d}t,\end{aligned} \tag{53}$$

where the scale of the wavelet relates to that of the window w as $2j-j_0\le j'\le 2j+j_0$.

As previously we will use the vanishing moments of the wavelet to bring out a few $2^{-j'}$ factors. This time we will need up to three vanishing moments, i.e., we write the wavelet as

$$\psi^{\varepsilon}\big(2^{j'}s-n_1,2^{j'}t-n_2\big)=\left(2^{-j'}\frac{\mathrm{d}}{\mathrm{d}s}\right)^M\tilde{\psi}^{\varepsilon}\big(2^{j'}s-n_1,2^{j'}t-n_2\big),\quad \varepsilon=\mathrm{HL\ or\ HH},$$

or as

$$\psi^{\varepsilon}\big(2^{j'}s-n_1,2^{j'}t-n_2\big)=\left(2^{-j'}\frac{\mathrm{d}}{\mathrm{d}t}\right)^M\tilde{\psi}^{\varepsilon}\big(2^{j'}s-n_1,2^{j'}t-n_2\big),\quad \varepsilon=\mathrm{LH},$$

where $M\le 3$, and $\tilde{\psi}^{\varepsilon}$ has the same support as ψ^{ε}. Before we let these derivatives act on the rest of the integrand, multiply and divide by $\phi(s,t)=\|\mathbf{x}(t)-\mathbf{x}(s)\|$ to

get

$$k H_1^{(1)}\big(k\phi(s,t)\big)\big(\dot{\mathbf{x}}(t)\big)^{\perp}\cdot r = \big[k\phi(s,t)H_1^{(1)}\big(k\phi(s,t)\big)\big]\big(\dot{\mathbf{x}}(t)\big)^{\perp}\cdot\frac{\mathbf{x}(t)-\mathbf{x}(s)}{\|\mathbf{x}(t)-\mathbf{x}(s)\|^2}.$$

We will then need the following lemma, which refines the results of (37), (38), and Lemma 6. It is proved in the Appendix.

Lemma 7 *Let $\phi(s,t)=\|\mathbf{x}(s)-\mathbf{x}(t)\|$. For every integer $m\geq 0$ there exists $C_m>0$ such that, as long as $s\neq t$,*

$$\left|\left(\frac{\mathrm{d}}{\mathrm{d}s}\right)^m\phi(s,t)\right|\leq C_m, \tag{54}$$

$$\left|\left(\frac{\mathrm{d}}{\mathrm{d}s}\right)^m\left[\big(\dot{\mathbf{x}}(t)\big)^{\perp}\cdot\frac{\mathbf{x}(t)-\mathbf{x}(s)}{\|\mathbf{x}(t)-\mathbf{x}(s)\|^2}\right]\right|\leq C_m. \tag{55}$$

The same inequalities hold with $\mathrm{d}/\mathrm{d}t$ *derivatives in place of* $\mathrm{d}/\mathrm{d}s$ *derivatives.*

Let us first consider the wavelets whose support intersects the diagonal. There, the support of the wavelet is sufficiently small that $\chi_j^{\mathrm{diag}}(s,t)=\rho(k\sigma)=1$. One may integrate by parts only once in (53), because

$$\nabla_{(s,t)}\phi(s,t)=\big(\dot{\mathbf{x}}(s)\cdot r,-\dot{\mathbf{x}}(t)\cdot r\big)$$

is discontinuous, since the unit chord $r=(\mathbf{x}(s)-\mathbf{x}(t))/\|\mathbf{x}(s)-\mathbf{x}(t)\|$ changes sign across the diagonal and $\dot{\mathbf{x}}(t)\neq 0$ there. The action of either $\mathrm{d}/\mathrm{d}s$ or $\mathrm{d}/\mathrm{d}t$ on the integrand after integration by parts gives:

- An $O(1)$ contribution for $\frac{\mathrm{d}}{\mathrm{d}x}(xH_1^{(1)}(x))$, by Lemma 2.
- By the chain rule, an $O(k)$ contribution for $\frac{\mathrm{d}}{\mathrm{d}s}(k\phi)$ and $\frac{\mathrm{d}}{\mathrm{d}t}(k\phi)$, because of Lemma 7.
- An $O(1)$ contribution for derivatives of $(\dot{\mathbf{x}}(t))^{\perp}\cdot r/\phi$, by Lemma 7.

The size of the support is $2^{-2j'}$, the wavelet comes with an L^2 normalization $2^{j'}$, one factor $2^{-j'}$ comes out of the vanishing moment, and $|j'-2j|\leq$ const.; hence diagonal wavelet coefficients obey the bound

$$\big|\big\langle K\chi_j^{\mathrm{diag}}\rho(k\sigma),\psi^{\varepsilon}_{j',\mathbf{n}}\big\rangle\big|\leq 2^{-4j}k\quad\text{(diagonal wavelets)}.$$

There are $O(2^{j'})=O(2^{2j})$ such diagonal wavelets overall, hence the sum of squares of these coefficients is bounded by $2^{-6j}k^2$, in accordance with (52).

Let us now treat the wavelets that do not intersect the diagonal $s=t$, and show that the same bound is valid. One will now need to integrate by parts three times in s or t to get three $2^{-j'}$ factors out, and gather the action of the derivatives on the rest of the integrand as follows.

- The factors $\rho(k\sigma)\chi_j^{\mathrm{diag}}(s,t)$ are essentially multiplied by $\max(2^j,k)$ for each derivative.
- By Lemma 2, the combination $xH_1^{(1)}(x)$ becomes $1/x$ when differentiated three times in x. This is $1/k\phi$ when $x=k\phi$.

– Derivatives of ϕ in s and t remain $O(1)$ by Lemma 7, hence each derivative of $k\phi$ yields a factor k.
– By Lemma 7, all derivatives of $(\dot{\mathbf{x}}(t))^{\perp} \cdot \frac{\mathbf{x}(t)-\mathbf{x}(s)}{\|\mathbf{x}(t)-\mathbf{x}(s)\|^2}$ remain bounded uniformly in s and t.

The product rule yields many terms but the overall sum is controlled by the behavior of the "extreme" terms identified above, hence a factor $\max(2^{3j}, k^3) + k^3/k\phi$ under the integral sign. Since the wavelet has support well away from the diagonal, we can proceed as previously and bound $\phi^{-1}(s,t)$ by $q^{-1}2^{j'}$ where q is an integer indexing the distance between the diagonal and the center of the wavelet. Again, the support of the wavelet has area $O(2^{-2j'})$ and j' is comparable to $2j$, hence we get a bound

$$\left|\left\langle K\chi_j^{\mathrm{diag}}\rho(k\sigma), \psi^{\varepsilon}_{j',\mathbf{n}}\right\rangle\right| \le 2^{-8j}\left[\max\left(2^{3j}, k^3\right) + k^2 q^{-1} 2^{2j}\right]$$
(non-diagonal wavelets).

As seen previously, the number of wavelet coefficients is $O(\min(2^{2j}/k, 2^j))$ across the diagonal (indexed by q), times $O(2^{2j})$ along the diagonal, for a total of $O(\min(2^{4j}/k, 2^{3j}))$. Hence we have

$$\begin{aligned}
&\sum_{j'\in[2j-j_0,2j+j_0]} \sum_{\varepsilon,\mathbf{n}} \left|\left\langle K_Q\rho(k\sigma), \psi^{\varepsilon}_{j',\mathbf{n}}\right\rangle\right|^2 \\
&\lesssim \left[\min\left(\frac{2^{4j}}{k}, 2^{3j}\right) \times 2^{-16j}\max\left(2^{6j}, k^6\right)\right] + \left[2^{2j}\sum_q 2^{-16j}k^4 q^{-2} 2^{4j}\right] \\
&\lesssim \left[2^{-12j}k^5 + 2^{-7j}\right] + \left[2^{-10j}k^4\right] \\
&\lesssim 2^{-6j}k^2 \quad \text{since } j \ge \frac{1}{2}\log_2 k + C.
\end{aligned}$$

This is the desired decay rate, compatible with (52).

We are now left with the task of verifying that (52) implies the correct decay of the ℓ_p norm, as in (21). Let $p < 1$. Start by using Hölder's inequality (47) with $N = O(2^{j(2+\delta)})$—there are $O(2^j(1+\delta))$ wave atoms per square Q, and $O(2^j)$ squares along the diagonal. We get

$$\sum_{\mathbf{m},\mathbf{n}} \left|\left\langle K\chi_j^{\mathrm{diag}}\rho(k\sigma), \varphi_{j,\mathbf{m},\mathbf{n}}\right\rangle\right|^p \le C\left[\left(2^{-3j}k\right)2^{j(2+\delta)(1/p-1/2)}\right]^p = Ck^p 2^{(2-4p+\delta' p)j},$$

where $\delta' = \delta(\frac{1}{p} - \frac{1}{2})$. Finally, the p-triangle inequality asks to sum this bound over $j \ge \frac{1}{2}\log_2 k$. The sum is convergent provided $p > 1/2$ and δ' is taken sufficiently small. The result is

$$\|K_\mu\|^p_{\ell_p(F)} \le C_p k^p k^{\frac{1}{2}(2-4p+\delta' p)} = C_p k^{1-p+\delta''}, \quad \forall \delta'' > 0$$

After taking the $1/p$-th power, we fall exactly into scenario 3 for the ℓ_p summation, i.e., (21). The proof is complete.

2.10 Proof of Corollary 2

Passing to relative error estimates requires scaling ϵ by $1/\sqrt{k}$ and $\sqrt{k}$ respectively. Recall $k \geq 1$.

- If we invoke Theorem 1 for G_0, with an absolute error $\epsilon/\sqrt{k}$ in place of ϵ, then the number of terms $|\Lambda_0|$ becomes $O((k\epsilon^{-2})^{1+1/\infty})$. Hence if we can show that $1/\sqrt{k} \leq C\|K^0\|_2$, then (15) follows.
- If we invoke Theorem 1 for G_1, with an absolute error $\epsilon\sqrt{k}$ in place of ϵ, then the number of terms $|\Lambda_1|$ becomes $O(k^{1+1/\infty}\epsilon^{-1/\infty} + (k\epsilon^{-2})^{1/3+1/\infty})$. Hence if we can show that $\sqrt{k} \leq C\|K^1\|_2$, then (16) follows.
- The combination of the above two error bounds would show (17), provided we can show that $\sqrt{k} \leq C\|K^1 - \mathrm{i}\eta K^0\|_2$ when $\eta \asymp k$.

Therefore, it suffices to establish the lower bounds on $\|K^0\|_2$, $\|K^1\|_2$, and $\|K^{(0,1)}\|_2$. By the tight-frame property of wave atoms, the claim for K^0 is exactly $\int_{[0,1]^2} |G_0(s,t)|^2 \,\mathrm{d}s\,\mathrm{d}t \geq C/k$. Set $\phi(s,t) = \|\mathbf{x}(s) - \mathbf{x}(t)\|$. As $k\phi(s,t) > c_0$, Lemma 3 implies that $|G_0(s,t)| \leq C(k\phi)^{-1/2}$. If k is sufficiently large, we can restrict the integration domain to the nonempty set $k\phi(s,t) > c_0$, and directly conclude. If k is not large enough for this step, the integral is still a uniformly continuous and positive function of k, hence uniformly bounded away from zero.

The claim for K^1 is $\int_{[0,1]^2} |G_1(s,t)|^2 \,\mathrm{d}s\,\mathrm{d}t \geq k$. The non-Hankel factors in the expression of G_1 play a minor role in evaluating this lower bound; we can consider them bounded away from zero on a large set S_1 to which the integral is restricted. Lemma 3 then implies that for $k\phi(s,t) > c_1$ and $(s,t) \in S_1$, we have $|G_1(s,t)| \geq k(k\phi)^{-1/2}$. By the same reasoning as previously, this leads to the lower bound.

The claim for $K^{(0,1)}$ is $\int_{[0,1]^2} |G_1(s,t) - \mathrm{i}\eta G_0(s,t)|^2 \,\mathrm{d}s\,\mathrm{d}t \geq k$, with $\eta \asymp k$. The reasoning is here a little more complicated since G_1 and kG_0 are on the same order of magnitude. The presence of $-\mathrm{i}$, however, prevents major cancellations—and is in fact chosen for that very reason. The asymptotic decay of $G_1 - \mathrm{i}\eta G_0$ for large $\phi(s,t)$ can be studied from the integral formulation of the Hankel function used throughout the appendix for proving the three lemmas in Sect. 2.1. Without entering into details, we remark that the integral factor in (56) is for z large very near real-valued, with positive real part. The exponential factor $\mathrm{e}^{-\mathrm{i}n\pi/2}$ shows that H_1 is then almost aligned with $-\mathrm{i}H_0$. The particular combination $G_1 - \mathrm{i}\eta G_0$ with $\eta > 0$ respects this quadrature property of Hankel functions, and produces no cancellation at all in the limit $z \to \infty$. So for $k\phi$ large enough, the claim follows; and if $k\phi$ is not large enough, we fall back on an argument of uniform continuity as previously.

3 Numerical Experiments

In this section, we provide several numerical examples to support the sparsity results of the previous section. The three geometric objects used in this section are displayed in Fig. 1. For each object, the boundary curve is represented using a small number of Fourier coefficients. The last two examples have non-convex shapes that typically result in multiple scattering effects. We report the numerical results for the single-layer kernel $G_0(s,t)$ in Sect. 3.1 and the results for the double-layer kernel $G_1(s,t)$

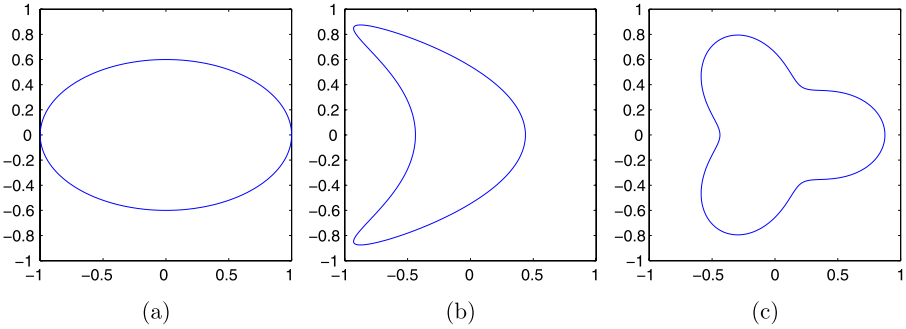


Fig. 1 The geometric objects used in the test examples. **a** An ellipse. **b** A kite-shaped object. **c** A star-shaped object

in Sect. 3.2. We omit the results of the combined kernel $G_1(s,t) - \mathrm{i}\eta G_0(s,t)$ as they are almost the same as the single-layer case.

3.1 Single-Layer Potential

We first study the single-layer potential

$$k \cdot G_0(s,t) = k \cdot \frac{\mathrm{i}}{4} H_0^{(1)}\big(k\big\|\mathbf{x}(s) - \mathbf{x}(t)\big\|\big)\big\|\dot{\mathbf{x}}(t)\big\|.$$

Notice that we use $k \cdot G_0(s,t)$ instead of $G_0(s,t)$, because the coupling constant η in the integral equation (1) is of order k. Therefore, $k \cdot G_0(s,t)$ is more informative when we report the value at which coefficients are thresholded, and the number of non-negligible coefficients.

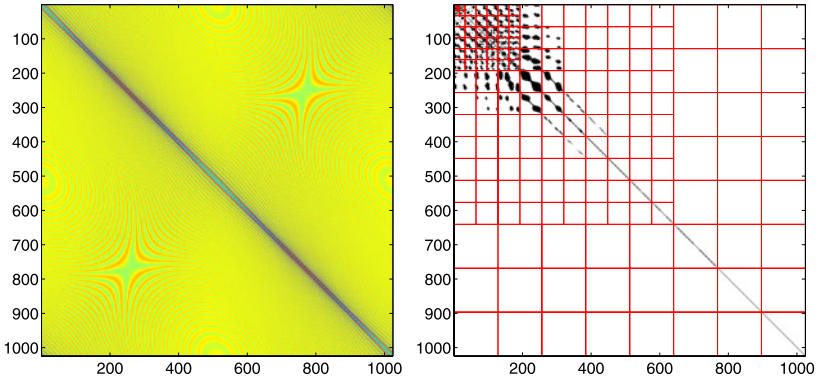
For each fixed k, we construct the discrete version of the operator $k \cdot G_0(s,t)$ by sampling the boundary curve with $N = 8k$ quadrature points; this corresponds to about eight points per wavelength $(2\pi/k)$ in these examples. Next, we scale the values at these quadrature points with the high-order corrected trapezoidal quadrature rule from [17] in order to integrate the logarithmic singularity accurately. This quadrature rule has the appealing feature of changing the weights only locally close to the singularity. We then apply the two-dimensional wave atom transform to compute the coefficients $K_\mu^0 := \langle kG_0, \varphi_\mu \rangle$. For a fixed accuracy ϵ, we obtain the sparsest approximant $\tilde{K}_\mu^0$ that satisfies

$$\big\|K^0 - \tilde{K}^0\big\|_{\ell_2(\mu)} \le \epsilon \big\|K^0\big\|_{\ell_2(\mu)}$$

by choosing the largest possible threshold value δ and setting the coefficients less than δ in modulus to zero. Equation (15) predicts that, as a function of k, the number of wave atom coefficients defining $\tilde{K}^0$ should grow like $k^{1+1/\infty}$.

For each example in Fig. 1, we perform the test for different combinations of (k,ϵ) with $k = 32, 64, \ldots, 1024$ and $\epsilon = 10^{-1}, 10^{-1.5}$, and 10^{-2}. The numerical results for the three examples are summarized in Tables 1, 2, and 3, respectively. In each table, we have the following.

Table 1 Single-layer potential for the ellipse. Top left: the real part of the operator for $k = 128$. Top right: the sparsity pattern of the operator under the wave atom basis for $k = 128$ and $\epsilon = 10^{-2}$. Each black pixel stands for a non-negligible coefficient. Bottom: for different combinations of k and ϵ, the maximum number of non-negligible entries per row $|\Delta_0|/N$, the threshold value δ, and the estimated L^2 operator norm ϵ_{L^2} (from left to right)

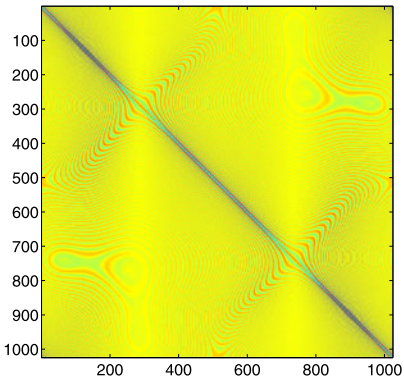


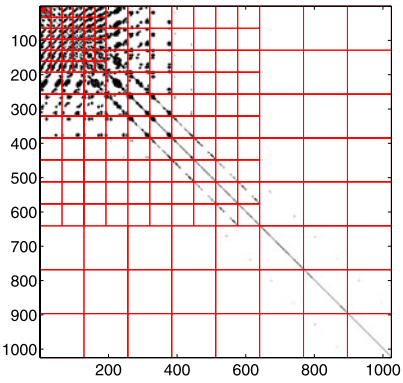
	$\epsilon = 10^{-1}$	$\epsilon = 10^{-1.5}$	$\epsilon = 10^{-2}$
$k = 32$	11/3.26e–2/9.53e–2	18/9.09e–3/2.88e–2	27/2.60e–3/7.93e–3
$k = 64$	9/3.37e–2/1.07e–1	16/9.20e–3/3.39e–2	28/2.39e–3/1.01e–2
$k = 128$	10/3.33e–2/9.80e–2	20/8.46e–3/3.29e–2	32/2.45e–3/1.04e–2
$k = 256$	9/3.63e–2/9.79e–2	18/9.33e–3/2.99e–2	30/2.46e–3/9.66e–3
$k = 512$	11/3.61e–2/9.79e–2	20/9.39e–3/3.07e–2	33/2.55e–3/9.86e–3
$k = 1024$	9/4.00e–2/9.78e–2	17/1.01e–2/3.12e–2	29/2.72e–3/9.81e–3

- The top left plot is the real part of the single-layer potential in the case of $k = 128$. This plot displays coherent oscillatory patterns for which the wave atom frame is well suited.
- The top right plot is the sparsity pattern of the operator under the wave atom basis for $k = 128$ and $\epsilon = 10^{-2}$. Each black pixel stands for a non-negligible coefficient. The coefficients are organized in a way similar to the usual ordering of 2D wave atom coefficients: each block contains the wave atom coefficients of a fixed frequency index $(j, \mathbf{m})$, and the blocks are ordered such that the lowest frequency is located at the top left corner while the highest frequency at the bottom right corner. Within a block, the wave atom coefficients of frequency index $(j, \mathbf{m})$ are ordered according to their spatial locations. The multiscale nature of the wave atom frame can be clearly seen from this plot.
- The table at the bottom gives, for different combinations of k and ϵ, the number of non-negligible coefficients per row $|\Delta_0|/N$, the threshold value δ (coefficients below this value in modulus are put to zero) and the L^2-to-L^2 norm operator error ϵ_{L^2} estimated using random test functions.

In these tables, the number of significant coefficients per row Δ_0/N grows very slowly as k doubles and reaches a constant level for large values of k. This matches well with the theoretical analysis in Sect. 2. The threshold value δ remains roughly at

Table 2 Single-layer potential for the kite-shaped object. Top left: the real part of the operator for $k = 128$. Top right: the sparsity pattern of the operator under the wave atom basis for $k = 128$ and $\epsilon = 10^{-2}$. Bottom: for different combinations of k and ϵ, $|\Delta_0|/N$, δ, and ϵ_{L^2}





	$\epsilon = 10^{-1}$	$\epsilon = 10^{-1.5}$	$\epsilon = 10^{-2}$
$k = 32$	14/2.92e–2/9.50e–2	25/8.35e–3/3.07e–2	37/2.69e–3/8.79e–3
$k = 64$	15/2.70e–2/9.77e–2	30/7.66e–3/3.45e–2	46/2.41e–3/9.98e–3
$k = 128$	17/2.68e–2/1.04e–1	34/7.49e–3/3.35e–2	53/2.26e–3/1.04e–2
$k = 256$	17/2.70e–2/1.06e–1	35/7.27e–3/3.36e–2	58/2.07e–3/1.05e–2
$k = 512$	17/2.85e–2/1.11e–1	35/7.58e–3/3.43e–2	58/2.09e–3/1.07e–2
$k = 1024$	17/2.89e–2/1.03e–1	35/7.75e–3/3.33e–2	60/2.07e–3/1.07e–2

a constant level as k grows, which is quite different from the results obtained using wavelet packet bases [11, 12, 15, 16] where the threshold value in general decreases as k grows. The estimated L^2-to-L^2 operator error ϵ_{L^2} is very close to the prescribed accuracy ϵ in all cases. This indicates that, in order to get an approximation within accuracy ϵ in operator norm, one can simply truncate the non-standard form of the operator in the wave atom frame with the same accuracy.

3.2 Double-Layer Potential

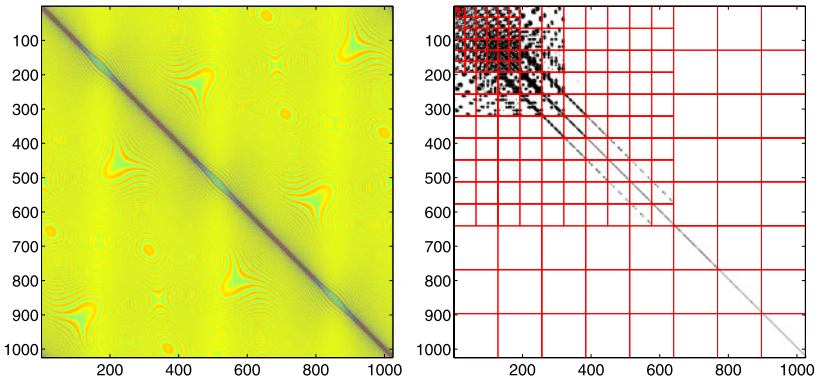
We now consider the double-layer potential

$$G_1(s,t) = \frac{\mathrm{i}k}{4} H_1^{(1)}\big(k\|\mathbf{x}(s) - \mathbf{x}(t)\|\big) \frac{\mathbf{x}(s) - \mathbf{x}(t)}{\|\mathbf{x}(s) - \mathbf{x}(t)\|} \cdot n_{\mathbf{x}(t)} \|\dot{\mathbf{x}}(t)\|.$$

For each fixed k, the discrete version of $G_1(s,t)$ is constructed by sampling at $N = 8k$ points and using trapezoidal quadrature rule. The coefficients $K_\mu^1 := \langle G_1, \varphi_\mu \rangle$ are calculated using the two-dimensional wave atom transform and the approximant $\tilde{K}_\mu^1$ is constructed in the same way as the single-layer potential case.

The results of the double-layer potentials for the three examples are summarized in Tables 4, 5, and 6, respectively. These results are qualitatively similar to the ones of the singular layer potential. However, the coefficients of the double-layer potential exhibit better sparsity pattern for the simple reason that the double-layer potential oper-

Table 3 Single-layer potential for the star-shaped object. Top left: the real part of the operator for $k = 128$. Top right: the sparsity pattern of the operator under the wave atom basis for $k = 128$ and $\epsilon = 10^{-2}$. Bottom: for different combinations of k and ϵ, $|\Delta_0|/N$, δ, and ϵ_{L^2}



	$\epsilon = 10^{-1}$	$\epsilon = 10^{-1.5}$	$\epsilon = 10^{-2}$
$k = 32$	14/2.76e–2/1.00e–1	23/8.74e–3/3.14e–2	32/2.50e–3/9.72e–3
$k = 64$	14/2.55e–2/8.97e–2	25/7.68e–3/2.83e–2	38/2.39e–3/8.28e–3
$k = 128$	18/2.39e–2/8.36e–2	33/7.01e–3/2.64e–2	49/2.19e–3/8.06e–3
$k = 256$	18/2.32e–2/9.48e–2	35/6.66e–3/2.90e–2	56/1.98e–3/9.33e–3
$k = 512$	19/2.34e–2/9.30e–2	40/6.19e–3/2.94e–2	66/1.84e–3/9.30e–3
$k = 1024$	18/2.42e–2/9.50e–2	38/6.41e–3/3.01e–2	66/1.77e–3/9.41e–3

ator has a singularity much weaker than logarithmic along the diagonal (where $s = t$) for objects with smooth boundary. Therefore, for a fixed accuracy ϵ, the number of wave atoms required along the diagonal for the double-layer potential is smaller than the number for the singular layer potential. This is clearly shown in the sparsity pattern plots in Tables 4, 5, and 6.

Acknowledgements The first author is partially supported by an NSF grant. The second author is partially supported by an NSF grant, a Sloan Research Fellowship, and a startup grant from the University of Texas at Austin.

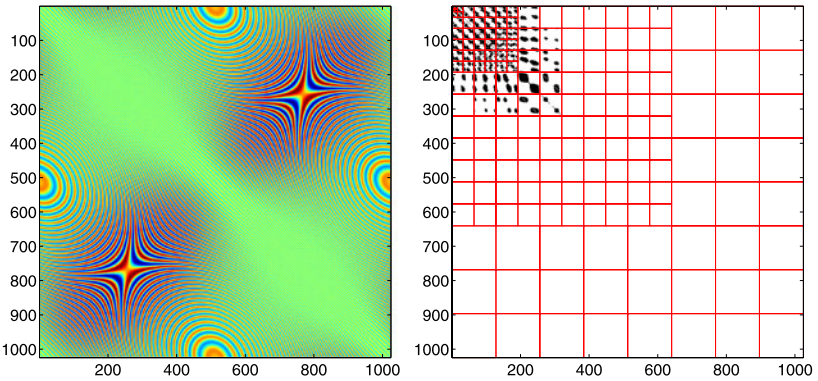
Appendix: Additional Proofs

Proof of Lemma 1 Following Watson's treatise [24], the Hankel function can be expressed by complex contour integration as

$$H_n^{(1)}(z) = \left(\frac{2}{\pi z}\right)^{1/2} \frac{\exp \mathrm{i}(z - \frac{n\pi}{2} - \frac{\pi}{4})}{\Gamma(n - \frac{1}{2})} \int_0^{\infty e^{\mathrm{i}\beta}} \mathrm{e}^{-u} u^{n-1/2} \left(1 + \frac{\mathrm{i}u}{2z}\right)^{n-1/2} \mathrm{d}u, \quad (56)$$

where $-\pi/2 < \beta < \pi/2$. For us, z is real and positive, and we take $\beta = 0$ for simplicity.

Table 4 Double-layer potential for the ellipse. Top left: the real part of the operator for $k = 128$. Top right: the sparsity pattern of the operator under the wave atom basis for $k = 128$ and $\epsilon = 10^{-2}$. Bottom: for different combinations of k and ϵ, $|\Delta_1|/N$, δ, and ϵ_{L^2}



	$\epsilon = 10^{-1}$	$\epsilon = 10^{-1.5}$	$\epsilon = 10^{-2}$
$k = 32$	10/1.39e–2/1.29e–1	16/4.61e–3/3.65e–2	21/1.57e–3/1.19e–2
$k = 64$	9/1.40e–2/1.00e–1	15/4.41e–3/3.23e–2	21/1.34e–3/1.01e–2
$k = 128$	11/1.21e–2/9.50e–2	20/3.74e–3/3.00e–2	28/1.12e–3/9.94e–3
$k = 256$	10/1.28e–2/9.80e–2	18/3.70e–3/3.28e–2	28/1.07e–3/1.02e–2
$k = 512$	13/1.16e–2/1.00e–1	22/3.42e–3/3.24e–2	33/9.77e–4/1.00e–2
$k = 1024$	11/1.21e–2/9.98e–2	20/3.52e–3/3.29e–2	30/9.83e–4/1.01e–2

Let us first treat the case $m = 0$ (no differentiations) and $n > 0$. We can use the simple bound

$$\left|1 + \frac{\mathrm{i}u}{2z}\right|^{n-\frac{1}{2}} \leq C_n\left(1 + \left|\frac{u}{z}\right|^{n-\frac{1}{2}}\right)$$

to see that the integral, in absolute value, is majorized by $C_n(1 + z^{-n+1/2})$. Hence the Hankel function itself is bounded by $C_n(z^{-1/2} + z^{-n})$. This establishes the first two expressions in (22) in the case $m = 0$.

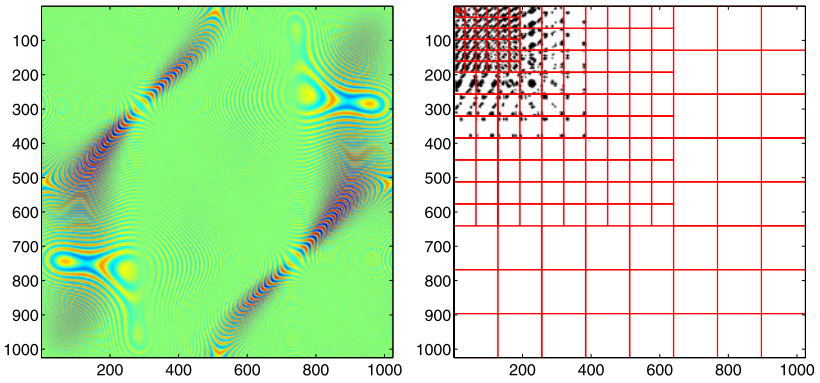
The case $m = n = 0$ is treated a little differently because the integrand in (56) develops a $1/u$ singularity near the origin as $z \to 0$. We have

$$\left|1 + \frac{\mathrm{i}u}{2z}\right|^{-\frac{1}{2}} = \left(1 + \left(\frac{u}{2z}\right)^2\right)^{-1/4} \leq C \min\left(1, \left(\frac{u}{z}\right)^{-1/2}\right), \tag{57}$$

hence the integral in (56) is bounded in modulus by a constant times

$$\int_0^z \mathrm{e}^{-u} u^{-1/2}\,\mathrm{d}u + z^{1/2}\int_z^\infty \mathrm{e}^{-u} u^{-1}\,\mathrm{d}u \leq C\left(z^{1/2} + z^{1/2}|\log z|\right).$$

Table 5 Double-layer potential for the kite-shaped object. Top left: the real part of the operator for $k = 128$. Top right: the sparsity pattern of the operator under the wave atom basis for $k = 128$ and $\epsilon = 10^{-2}$. Bottom: for different combinations of k and ϵ, $|\Delta_1|/N$, δ, and ϵ_{L^2}



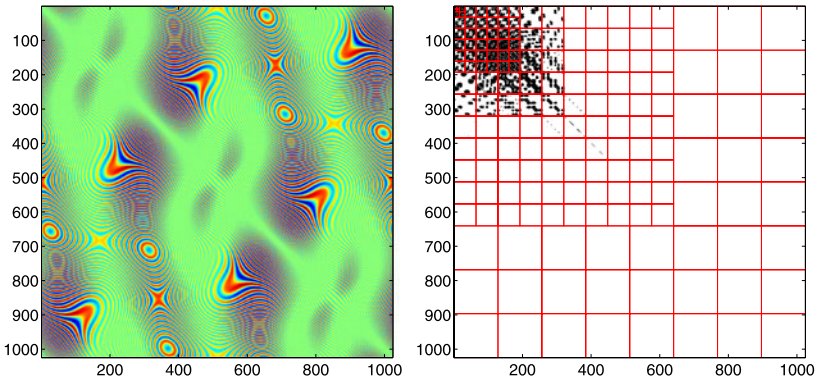
	$\epsilon = 10^{-1}$	$\epsilon = 10^{-1.5}$	$\epsilon = 10^{-2}$
$k = 32$	12/1.64e–2/9.16e–2	19/5.16e–3/2.82e–2	27/1.64e–3/8.71e–3
$k = 64$	13/1.40e–2/1.02e–1	23/4.37e–3/3.18e–2	34/1.31e–3/1.10e–2
$k = 128$	16/1.27e–2/9.11e–2	29/3.82e–3/2.80e–2	43/1.18e–3/9.27e–3
$k = 256$	16/1.22e–2/9.56e–2	31/3.46e–3/3.23e–2	49/1.02e–3/1.01e–2
$k = 512$	18/1.20e–2/9.98e–2	33/3.42e–3/3.09e–2	52/9.71e–4/9.81e–3
$k = 1024$	18/1.17e–2/9.57e–2	35/3.26e–3/3.02e–2	55/9.07e–4/1.03e–2

With the $z^{-1/2}$ factor from (56), the resulting bound is $C(1+|\log z|)$ as desired (third equation). When $z > 1$, we can improve this to $C(1+z^{1/2}\mathrm{e}^{-z}) \le C$, which gives the first equation when $m = n = 0$.

For the case $m > 0$, it suffices to apply Leibniz's rule inductively and observe that each derivative produces a factor x^{-1} without changing the power of k. In particular, we have the following.

- With $\alpha \neq 0$,
$$\frac{\mathrm{d}}{\mathrm{d}x}\left[(kx)^{-\alpha}\right] = -\alpha(kx)^{-\alpha}\frac{1}{x},$$
hence the power of k is preserved and one negative power of x is created.
- Derivatives acting on isolated negative powers of x also produce an x^{-1} factor without affecting the dependence on k.
- As for the x dependence under the integral sign, with $\alpha \neq 0$, we arrange the factors as
$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\left(1+\frac{\mathrm{i}u}{2kx}\right)^{-\alpha}\right) = \alpha\left[\frac{\mathrm{i}u}{2kx}\left(1+\frac{\mathrm{i}u}{2kx}\right)^{-1}\right]\left(1+\frac{\mathrm{i}u}{2kx}\right)^{-\alpha}\frac{1}{x}.$$
The factor in square brackets is bounded by 1 in modulus, hence the dependence on k is unchanged. The factor $1/x$ is the only modification in the dependence on x. Subsequent differentiations will only act on factors that we have already treated above: powers of kx, powers of x, and powers of $1+\frac{\mathrm{i}u}{2kx}$. This finishes the proof.

Table 6 Double-layer potential for the star-shaped object. Top left: the real part of the operator for $k = 128$. Top right: the sparsity pattern of the operator under the wave atom basis for $k = 128$ and $\epsilon = 10^{-2}$. Bottom: for different combinations of k and ϵ, $|\Delta_1|/N$, δ, and ϵ_{L^2}



	$\epsilon = 10^{-1}$	$\epsilon = 10^{-1.5}$	$\epsilon = 10^{-2}$
$k = 32$	13/1.43e–2/1.20e–1	20/5.15e–3/3.31e–2	25/1.62e–3/1.15e–2
$k = 64$	14/1.25e–2/9.20e–2	23/4.23e–3/2.80e–2	31/1.43e–3/9.03e–3
$k = 128$	19/1.08e–2/1.08e–1	32/3.42e–3/3.27e–2	44/1.19e–3/1.10e–2
$k = 256$	20/1.00e–2/1.02e–1	36/3.08e–3/3.13e–2	52/9.77e–4/1.07e–2
$k = 512$	23/8.99e–3/1.01e–1	44/2.66e–3/3.24e–2	66/8.22e–4/1.01e–2
$k = 1024$	22/9.06e–3/1.04e–1	43/2.54e–3/3.21e–2	69/7.46e–4/9.46e–3

Proof of Lemma 2 As previously, we use the integral formulation to get

$$x H_1^{(1)}(x) = f(x) \int_0^\infty \mathrm{e}^{-u} u^{1/2} \left(x + \mathrm{i}\frac{u}{2} \right)^{1/2} \mathrm{d}u,$$

where $f(x)$ is the exponential factor, and already obeys $|f^{(n)}(x)| \le C$ for all $n \ge 0$. We denote the integral factor by $I(x)$; its derivatives are

$$I^{(n)}(x) = C_n \int_0^\infty \mathrm{e}^{-u} u^{1/2} \left(x + \mathrm{i}\frac{u}{2} \right)^{\frac{1}{2}-n} \mathrm{d}u,$$

where C_n is a numerical constant. In a manner analogous to the proof of Lemma 1, we can bound

$$\left| x + \mathrm{i}\frac{u}{2} \right|^{\frac{1}{2}-n} = \left(x^2 + \frac{u^2}{4} \right)^{\frac{1}{4}-\frac{n}{2}} \le C_n \big(\max(x,u) \big)^{\frac{1}{2}-n}.$$

It follows that

$$|I^{(n)}(x)| \le C_n \left[\int_0^x \mathrm{e}^{-u} u^{1/2} x^{1/2-n} \, \mathrm{d}u + \int_x^\infty \mathrm{e}^{-u} u^{1-n} \, \mathrm{d}u \right].$$

In the first term we can use $e^{-u} \le 1$ and bound the integral by a constant times x^{2-n}. The integrand of the second term has a singularity near $u = 0$ that becomes more severe as n increases; this term is bounded by $O(1)$ if $n = 0$ or 1, by $O(1 + |\log(x)|)$ if $n = 2$, and by $O(x^{2-n})$ if $n > 2$.

Proof of Lemma 3 Consider (56) again, and take z real. For large values of z, the factor $(1 + iu/2z)^{n-1/2}$ is close to one; more precisely, it is easy to show that for each $n \ge 0$, there exists $c_n > 0, d_n > 0$ for which

$$\left| \left(1 + \frac{iu}{2z}\right)^{n-1/2} - 1 \right| \le d_n \frac{u}{z}, \quad \text{if } u \le c_n z.$$

We can insert this estimate in (56) and split the integral into two parts to obtain

$$\begin{aligned} \left|H_n^{(1)}(z)\right| \Gamma(n - 1/2) \left(\frac{\pi z}{2}\right)^{1/2} - \left| \int_0^{c_n z} e^{-u} u^{-1/2} \, du \right| \\ \ge -d_n \int_0^{z c_n} e^{-u} u^{-1/2} \frac{u}{z} \, du - C \int_{z c_n}^{\infty} e^{-u} u^{-1/2} \, du. \end{aligned}$$

The constant C in the last term comes from (57). The first term in the right-hand side is $O(z^{-1})$, and the second term is $O(e^{-z})$. At the expense of possibly choosing increasing the value of c_n, the second term in the left-hand side can manifestly be made to dominate the contribution of the right-hand side, proving the lemma.

Proof of Lemma 6 We start by writing

$$\big(\dot{\mathbf{x}}(t)\big)^{\perp} \cdot r = \big(\dot{\mathbf{x}}(t)\big)^{\perp} \cdot \frac{\mathbf{x}(t) + \sigma \dot{\mathbf{x}}(t) - \mathbf{x}(s)}{\|\mathbf{x}(t) - \mathbf{x}(s)\|}.$$

By Taylor's theorem, $|\mathbf{x}(t) + \sigma \dot{\mathbf{x}}(t) - \mathbf{x}(s)| \le C\sigma^2$, hence $|(\dot{\mathbf{x}}(t))^{\perp} \cdot r| \le C\sigma$. Derivatives are then treated by induction; recall that $\frac{d}{d\tau} = \frac{d}{ds} + \frac{d}{dt}$ and $\frac{d}{d\sigma} = \frac{d}{ds} - \frac{d}{dt}$;

- Any number of τ or σ derivatives acting on $(\dot{\mathbf{x}}(t))^{\perp}$ leave it $O(1)$.
- τ derivatives acting on $\mathbf{x}(t) + \sigma \dot{\mathbf{x}}(t) - \mathbf{x}(s)$ leave it $O(\sigma^2)$ while each σ derivative removes an order of σ.
- τ derivatives acting on $\|\mathbf{x}(t) - \mathbf{x}(s)\|^{-m}$ leave it $O(\sigma^{-m})$ (m is generic) while each σ derivative removes an order of σ.

This shows (50) and (51).

Proof of Lemma 7 Some cancellations will need to be quantified in this proof that were not a concern in the justification of previous coarser estimates like Lemma 6.

Without loss of generality, assume that $\mathbf{x}(t) = (0, 0)$, $n_{\mathbf{x}(t)} = (0, 1)$, and that we have performed a change of variables such that the curve is parametrized as the graph $\mathbf{x}(s) = (s, f(s))$ of some function $f \in C^{\infty}$ obeying $|f(s)| \le Cs^2$. This latter change of variables would contribute a bounded multiplicative factor that would not compromise the overall estimate.

By symmetry, if (54) is true for $\mathrm{d}/\mathrm{d}s$ derivatives, then it will be true for $\mathrm{d}/\mathrm{d}t$ derivatives as well. Without loss of generality let $s > 0$. Then we have

$$\phi(s,t) = \big\|\mathbf{x}(s) - \mathbf{x}(t)\big\| = \sqrt{s^2 + f^2(s)} = s\sqrt{1 + \frac{f^2(s)}{s^2}}.$$

Since $|f(s)| \le Cs^2$ and C^∞, the ratio $f^2(s)/s^2$ is also bounded for $s \lesssim 1$, and of class C^∞. Being a composition of C^∞ functions, the whole factor $\sqrt{1 + f^2/s^2}$ is therefore also of class C^∞, which proves (54).

As for (55) with $\mathrm{d}/\mathrm{d}s$ derivatives, we can write $(\dot{\mathbf{x}}(t))^\perp = (0,1)$ and

$$\big(\dot{\mathbf{x}}(t)\big)^\perp \cdot \frac{\mathbf{x}(t) - \mathbf{x}(s)}{\|\mathbf{x}(t) - \mathbf{x}(s)\|^2} = \frac{-f(s)}{s^2 + f^2(s)} = -\left(\frac{f(s)}{s^2}\right)\frac{1}{1 + \frac{f^2(s)}{s^2}}.$$

(The $1/\|\dot{\mathbf{x}}(t)\|$ does not pose a problem since it is C^∞.) Again, since $|f(s)| \le Cs^2$, both ratios $f(s)/s^2$ and $f^2(s)/s^2$ are themselves bounded and of class C^∞. The factor $\frac{1}{1+\frac{f^2(s)}{s^2}}$ is the composition of two C^∞ functions, hence also of class C^∞.

The symmetry argument is not entirely straightforward for justifying (55) with $\mathrm{d}/\mathrm{d}t$ derivatives. Symmetry $s \leftrightarrow t$ only allows us to conclude that

$$\left|\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^m \left[\big(\dot{\mathbf{x}}(s)\big)^\perp \cdot \frac{\mathbf{x}(t) - \mathbf{x}(s)}{\|\mathbf{x}(t) - \mathbf{x}(s)\|^2}\right]\right| \le C_m,$$

where $(\dot{\mathbf{x}}(s))^\perp$ appears in place of the desired $(\dot{\mathbf{x}}(t))^\perp$. Hence it suffices to show that the $\mathrm{d}/\mathrm{d}t$, or equivalently the $\mathrm{d}/\mathrm{d}s$ derivatives of

$$\big[\big(\dot{\mathbf{x}}(t)\big)^\perp - \big(\dot{\mathbf{x}}(s)\big)^\perp\big] \cdot \frac{\mathbf{x}(t) - \mathbf{x}(s)}{\|\mathbf{x}(t) - \mathbf{x}(s)\|^2}$$

stay bounded. Using our frame in which the curve is a graph, we find

$$\big(\dot{\mathbf{x}}(t)\big)^\perp - \big(\dot{\mathbf{x}}(s)\big)^\perp = (0,1) - \big(-f'(s), 1\big) = \big(f'(s), 0\big).$$

As a result,

$$\big[\big(\dot{\mathbf{x}}(t)\big)^\perp - \big(\dot{\mathbf{x}}(s)\big)^\perp\big] \cdot \frac{\mathbf{x}(t) - \mathbf{x}(s)}{\|\mathbf{x}(t) - \mathbf{x}(s)\|^2} = \frac{-sf'(s)}{s^2 + f^2(s)} = -\left(\frac{f'(s)}{s}\right)\frac{1}{1 + \frac{f^2(s)}{s^2}}.$$

Since $|f'(s)| \le s$, we are again in presence of a combination of C^∞ functions that stays infinitely differentiable.

References

1. E. Abbott, *Flatland: A Romance in Many Dimensions*, 3rd edn. (Dover, New York, 1992) (1884).
2. J.P. Antoine, R. Murenzi, Two-dimensional directional wavelets and the scale-angle representation, *Signal Process.* **52**, 259–281 (1996).

3. A. Averbuch, E. Braverman, R. Coifman, M. Israeli, A. Sidi, Efficient computation of oscillatory integrals via adaptive multiscale local Fourier bases, *Appl. Comput. Harmon. Anal.* **9**(1), 19–53 (2000).
4. G. Beylkin, R. Coifman, V. Rokhlin, Fast wavelet transforms and numerical algorithms. I, *Commun. Pure Appl. Math.* **44**(2), 141–183 (1991).
5. B. Bradie, R. Coifman, A. Grossmann, Fast numerical computations of oscillatory integrals related to acoustic scattering, *Appl. Comput. Harmon. Anal.* **1**, 94–99 (1993).
6. E.J. Candès, D.L. Donoho, New tight frames of curvelets and optimal representations of objects with piecewise-C^2 singularities, *Commun. Pure Appl. Math.* **57**, 219–266 (2004).
7. H. Cheng, W.Y. Crutchfield, Z. Gimbutas, L.F. Greengard, J.F. Ethridge, J. Huang, V. Rokhlin, N. Yarvin, J. Zhao, A wideband fast multipole method for the Helmholtz equation in three dimensions, *J. Comput. Phys.* **216**, 300–325 (2006).
8. A. Córdoba, C. Fefferman, Wave packets and Fourier integral operators, *Commun. PDE* **3**(11), 979–1005 (1978).
9. L. Demanet, L. Ying, Curvelets and wave atoms for mirror-extended images, in *Proc. SPIE Wavelets XII Conf.*, 2007.
10. L. Demanet, L. Ying, Wave atoms and sparsity of oscillatory patterns, *Appl. Comput. Harmon. Anal.* **23**(3), 368–387 (2007).
11. H. Deng, H. Ling, Fast solution of electromagnetic integral equations using adaptive wavelet packet transform, *IEEE Trans. Antennas Propag.* **47**(4), 674–682 (1999).
12. H. Deng, H. Ling, On a class of predefined wavelet packet bases for efficient representation of electromagnetic integral equations, *IEEE Trans. Antennas Propag.* **47**(12), 1772–1779 (1999).
13. B. Engquist, L. Ying, Fast directional multilevel algorithms for oscillatory kernels, *SIAM J. Sci. Comput.* **29**(4), 1710–1737 (2007).
14. B. Engquist, L. Ying, Fast directional computation for the high frequency Helmholtz kernel in two dimensions, *Commun. Math. Sci.* **7**(2), 327–345 (2009).
15. W.L. Golik, Wavelet packets for fast solution of electromagnetic integral equations, *IEEE Trans. Antennas Propag.* **46**(5), 618–624 (1998).
16. D. Huybrechs, S. Vandewalle, A two-dimensional wavelet-packet transform for matrix compression of integral equations with highly oscillatory kernel, *J. Comput. Appl. Math.* **197**(1), 218–232 (2006).
17. S. Kapur, V. Rokhlin, High-order corrected trapezoidal quadrature rules for singular functions, *SIAM J. Numer. Anal.* **34**, 1331–1356 (1997).
18. R. Kress, Minimizing the condition number of boundary integral operators in acoustic and electromagnetic scattering, *Q. J. Mech. Appl. Math.* **38**(2), 323–341 (1985).
19. S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd edn. (Academic Press, Orlando-San Diego, 1999).
20. F.G. Meyer, R.R. Coifman, Brushlets: a tool for directional image analysis and image compression, *Appl. Comput. Harmon. Anal.* **4**, 147–187 (1997).
21. V. Rokhlin, Rapid solution of integral equations of scattering theory in two dimensions, *J. Comput. Phys.* **86**(2), 414–439 (1990).
22. V. Rokhlin, Diagonal forms of translation operators for the Helmholtz equation in three dimensions, *Appl. Comput. Harmon. Anal.* **1**, 82–93 (1993).
23. L. Villemoes, Wavelet packets with uniform time-frequency localization, *C. R. Math.* **335**(10), 793–796 (2002).
24. G.N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd edn. (Cambridge University Press, Cambridge, 1966).