ON THE COMPRESSIVE SPECTRAL METHOD

ALAN MACKEY, HAYDEN SCHAEFFER, AND STANLEY OSHER

Abstract. The authors of [13] proposed sparse Fourier domain approximation of solutions to multiscale PDE problems by soft thresholding. We show here that the method enjoys a number of desirable numerical and analytic properties, including convergence for linear PDE and a modified equation resulting from the sparse approximation. We also extend the method to solve elliptic equations and introduce sparse approximation of differential operators in the Fourier domain. The effectiveness of the method is demonstrated on homogenization examples, where its complexity is dependent only on the sparsity of the problem and constant in many cases.

Key words. Spectral method, Compressed sensing, Sparse representation, Partial differential equations, Multiscale methods.

AMS subject classifications. 65N35, 65K10 , 65F50 , 65F10

1 Introduction Partial differential equations with multiple length scales are fundamental to modeling various physical problems including composite materials, wave propagation in inhomogeneous media, crystalline solids, and flows with high Reynolds number (fluid mechanics). Typically, these problems involve a wide range of scales, with each scale corresponding to a level of physical processes. However, in some cases, the problem is scale separable, in the sense that the mathematical representation of the dynamics involve one fine scale and one course scale. Even in this case, accurate numerical methods for solving these PDE can be computationally expensive since resolving both the coarse and fine scales simultaneously requires a spatial resolution dominated by the fine scale.

Over the past decades, various approaches have been taken to overcome this difficulty. In some cases, it is possible to derive an asymptotic approximation for the effect of small scales on the solution [11]. When this is not possible, many other techniques have been proposed. A multiscale finite elements method can be used to solve linear elliptic homogenization equations (see [8]), and has found many applications to other multiscale problems. The equation-free methods use accurate small scale and short time solvers to capture fine scale behavior and use them to govern the related course scale behavior [9]. The heterogeneous multiscale method [14] is a general numerical approache which also uses the scale separation of the problem to generate solvers on the micro and macroscopic levels. In [10], a projection based approach is used to construct an adaptive multiscale algorithm for elliptic homogenization equations. And more recently, a sparse transform method [5] exploits the scale separability of linear homogenization problems to construct a fast direct solver. The body of literature on multiscale models is large, and we only mention some of the popular methods. For more detail on general numerical methods for multiscale problems see [29, 14] and the citations therein.

In this work, we will focus our attention on linear partial differential equations with multiscale behavior either in the medium or in a source term. Following the work of [13], which used an L^1 optimization method to compress the Fourier coefficients of the solution, we build efficient solvers for periodic multiscale problems. In particular, we will use the sparse Fourier structure of solutions to construct numerical methods which solve the problem directly, without the separating the micro and macro scales explicitly.

L^1 optimization and its related models are at the center of many problems in the

fields of imaging science and data analysis, see for example [2, 17, 18, 16]. Due to the connection with sparse models for compressive sensing, recent works have introduced L^1 techniques for numerical partial differential equations. For example, in [13] L^1 regularized least squares was used to sparsely approximate the Fourier coefficients in multiscale dynamic PDE (and in this work we expand that approach). In [23, 24, 25], eigenfunctions with compact support were constructed to efficiently solve problems in quantum mechanics. Also, in [26] an L^1 nonlinear least squares model was used to sparsely recover coefficients of a second order ODE which are related to constructing intrinsic mode functions. In [27], low-rank libraries are used to sparsely approximate solutions to dynamical systems and thereby identify bifurcation regimes. Some theoretical results are provided in [21] for PDE with L^1-terms, related to some of these models. For more detailed analytic results, see [28, 1, 30] which laid the theoretical groundwork for these equations.

In this paper, we continue the work of [13] to leverage the sparsity of solutions in order to design an efficient numerical scheme. However, we also impose sparsity of the update operator to improve the complexity while retaining a similar level of accuracy. We show some theoretical results for the sparse spectral scheme and sparse operator-sparse solution spectral scheme. In particular, we provide error bounds between the solution and the sparse approximation as well as complexity bounds on the algorithm. Also, we continue to make connections between L^1 based methods and multiscale problems through a denoising interpretation of the homogenization expansion of the solution.

The outline of this work is as follows. In Section 2, we recall the explicit scheme from [13] and in 3 propose an implicit version as well as a sparse operator approximation. Theoretical results are provided in Section 4. A discussion on well-posedness is given in Section 5 and a denoising interpretation of the method is given in Section 6. In Section 7, some algorithmic analysis is provided. The algorithm is tested on numerical examples in Section 8, with concluding remarks given in Section 9.

1.1 Notation

- a – (or A for anisotropic problems) the medium inhomogeneity. $\hat{a}'$ is the sparse approximation of $\hat{a}$.
- μ – the shrink size variable. μ' is the corresponding variable for sparse operator approximation.
- k – the Fourier space variable, with positive and negative frequencies.
- Q – either a general numerical scheme or the matrix corresponding to a one-step linear numerical scheme.
- L – an elliptic operator. $\hat{L}$ is the operator when applied in the Fourier domain and $\hat{L}_h$ is its discretization. $\hat{L}'_h$ is the sparse approximation.

2 Preliminary We will consider linear multiscale problems where the solutions are sparse in the Fourier domain [5, 13]. For example, consider the parabolic problem:

$$\begin{aligned} \frac{\partial u^\epsilon}{\partial t} - \frac{\partial}{\partial x}\left(a(x/\epsilon)\frac{\partial u^\epsilon}{\partial x}\right) = 0 \quad &\text{on } [0, 2\pi] \text{ periodic} \\ u^\epsilon(x,0) = u_0^\epsilon(x), \quad &a(x/\epsilon) \text{ oscillatory.} \end{aligned} \tag{2.1}$$

Figure 2.1 shows the solution in physical and Fourier space. This phenomenon is common in multiscale PDE: distinct length scales manifest strikingly as sparsity in the frequency domain.

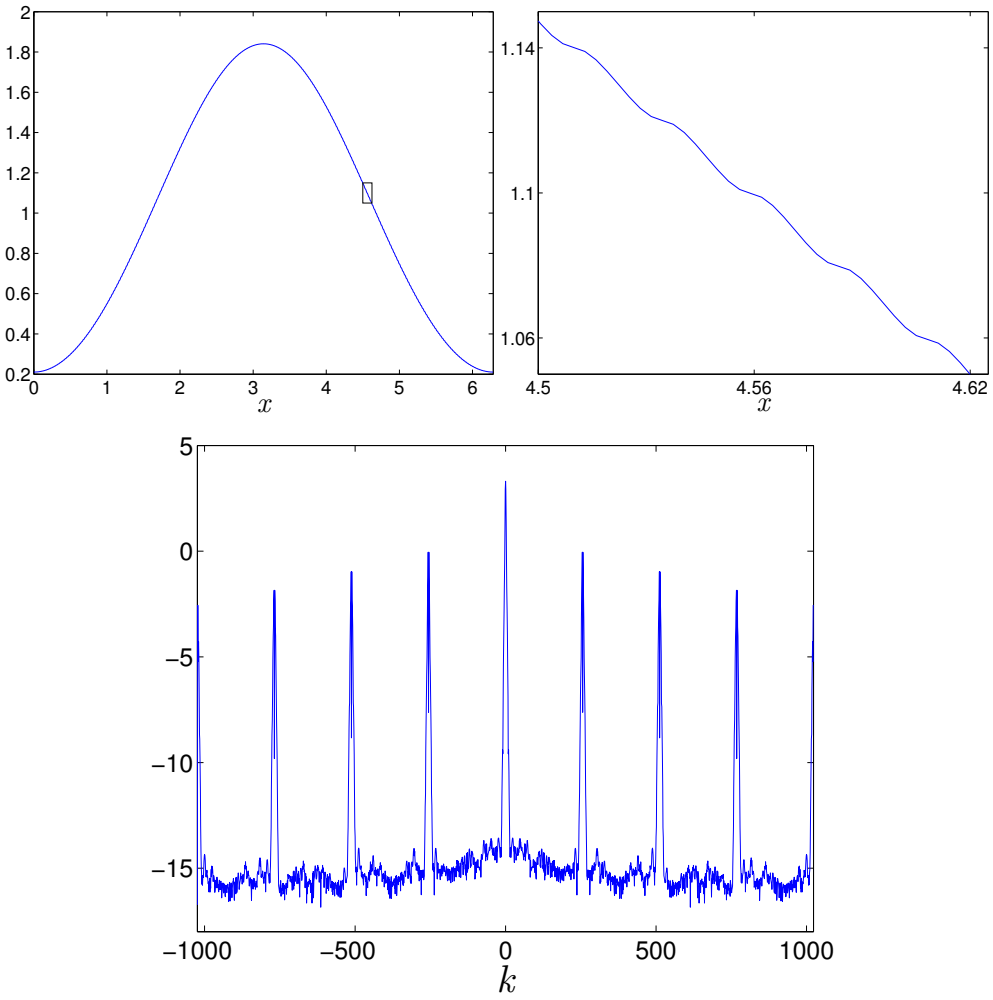

Fig. 2.1: **Left:** Solution of (2.1) with Fourier-sparse initial data in physical space. The small rectangle shows the axis limits of the zoomed in plot to the right. **Right:** Zoomed in, showing fine scale oscillations. **Bottom:** solutions in Fourier space (the y-axis for all Fourier space plots is on a $\log_{10}$ scale). Of the $N = 2048$ Fourier coefficients, only 153 have magnitude larger than 10^{-10}.

To compute solutions which are truly sparse in the frequency domain (and not just approximately sparse with many noisy small magnitude coefficients), it was proposed in [13] to solve an ℓ^1-regularized least squares problem to obtain a sparse approximation of $\hat{u}$ (the Fourier transform of u). We summarize the method here.

Given numerical iterates $\hat{u}^n, \ldots, \hat{u}^{n-q}$ and a numerical update scheme of the form

$$\hat{u}^{n+1} = Q(\hat{u}^n, \ldots, \hat{u}^{n-q}),$$

the scheme is modified by defining the auxiliary variable $\hat{v} = Q(\hat{u}^n, \ldots, \hat{u}^{n-q})$ and

solving

$$\hat{u}^{n+1} = \underset{w}{\operatorname{argmin}} \ \mu\|w\|_1 + \frac{1}{2}\|w - \hat{v}\|_2^2, \tag{2.2}$$

where the ℓ^1 norm for complex arguments w is $\|w\|_1 = \sum_i |w_i|$, where $|\cdot|$ denotes magnitude. Note that the ℓ^1 norm is taken in the Fourier domain and not physical space.

For a one-step linear updating scheme, equation (2.2) can be written as

$$\hat{u}^{n+1} = \underset{w}{\operatorname{argmin}} \ \mu\|w\|_1 + \frac{1}{2}\|w - Q\hat{u}^n\|_2^2$$

where Q is the matrix which advances the discretized solution forward in time. L^1 regularized least squares is amenable to a number of efficient solution methods, e.g. [19]. The problem can also be generalized to any basis or overcomplete dictionary, but we restrict our attention to Fourier modes. In fact, due to the orthogonality of the Fourier modes, equation (2.2) decouples and the minimizer can be given exactly:

$$\hat{u}^{n+1} = \operatorname{shrink}(\hat{v}, \mu) := \max(|\hat{v}| - \mu, 0)\frac{\hat{v}}{|\hat{v}|}.$$

For a concrete example, the forward Euler method applied to $\frac{\partial u^\epsilon}{\partial t} = \frac{\partial}{\partial x}\left[a(x/\epsilon)\frac{\partial u^\epsilon}{\partial x}\right]$ has the form:

$$\hat{u}^{n+1} = \hat{u}^n + dt\, i\, k\, \hat{a} * (i\, k\, \hat{u}^n)$$

where k is the wave number and $*$ represents convolution. This becomes:

$$\hat{u}^{n+1} = \operatorname{shrink}\left(\, \hat{u}^n + dt\, i\, k\, \hat{a} * (i\, k\, \hat{u}^n), \mu \,\right)$$

in the sparse spectral form. By exploiting sparsity in the frequency domain, the proposed method can rely on sparse data structures to allow for high resolution with faster numerical simulations.

3 Proposed Methods In this section, we will discuss two new extensions of the sparse spectral method, namely an implicit version and a sparse operator/sparse solution version. Each come with their own advantages, which we will analyze in subsequent sections.

3.1 Implicit Variation For many classes of problems at high spatial resolution, explicit schemes are impractical due to the severe time step restriction required for stability. We can construct an implicit scheme for the problems we are considering, which avoids these restrictions at the expense solving a more complex L^1 problem at each time step.

Consider the general linear parabolic equation $u_t + Lu = f$ with schemes of the form

$$Q\hat{u}^{n+1} = \hat{u}^n + dt\hat{f}_h.$$

The simplest implicit version is backward Euler:

$$(I + dt\hat{L}_h)\hat{u}^{n+1} = \hat{u}^n + dt\hat{f}_h,$$

where $\hat{L}$ denotes the representation of L in the Fourier basis, $\hat{L}_h$ denotes the discretized version of this operator with respect to a grid size $h > 0$, and $\hat{f}_h$ denotes the Fourier transform of f sampled at the corresponding grid points.

For a scheme of this form, the analogue of equation (2.2) is

$$\hat{u}^{n+1} = \underset{w}{\operatorname{argmin}} \ \mu\|w\|_1 + \frac{1}{2}\|Qw - (\hat{u}^n + dt\hat{f}_h)\|_2^2 \tag{3.1}$$

which does not have a simple explicit representation. In addition, the optimality condition for Equation (3.1) requires inverting the matrix Q^TQ which will often be badly conditioned.

When L is a uniformly elliptic operator, the eigenvalues of $Q = I + dt\hat{L}_h$ are positive and so we can instead consider the sparse scheme defined by

$$\hat{u}^{n+1} = \underset{w}{\operatorname{argmin}} \ \mu\|w\|_1 + \frac{1}{2}w^TQw - w^T(\hat{u}^n + dt\hat{f}_h). \tag{3.2}$$

Similarly for time-independent problems, *i.e.* $Lu = f$, the corresponding energy is

$$\hat{u} = \underset{w}{\operatorname{argmin}} \ \mu\|w\|_1 + \frac{1}{2}w^T\hat{L}_hw - w^T\hat{f}_h.$$

Note that when $\mu = 0$, this is the standard variational principle for elliptic operators. We will see that solving the implicit schemes with the L^1 term directly is often too slow to be practical. The reason is that directly applying this variational principle to find the solution does not use the fact that the solution is sparse in order to speed up computations. However, in Section 7.1 we will show that it is possible to construct an efficient algorithm for approximately solving the resulting optimality condition arising from equation (3.2).

3.2 Sparse Operator Approximation For uniformly elliptic linear operators, for example of the form

$$Lu = -\operatorname{div}(a(x, x/\epsilon)\nabla u),$$

the standard spectral discretization

$$\hat{L}_h\hat{u} = k\,\hat{a} * (k\,\hat{u})$$

requires a convolution at each iteration, which can be costly even when $\hat{u}$ is sparse. However, because the diffusion coefficient a is scale separated, we can define a sparse approximation of $\hat{L}_h$ by

$$\hat{L}'_h\hat{u} = k\,\hat{a}' * (k\,\hat{u})$$

where $\hat{a}'$ is a sparse approximation of $\hat{a}$. We can choose $\hat{a}'$ to solve

$$\hat{a}' = \underset{w}{\operatorname{argmin}} \ \mu'\|w\|_1 + \frac{1}{2}\|w - \hat{a}\|_2^2$$

which again results in a closed form solution given by the soft thresholding $\hat{a}' = \operatorname{shrink}(\hat{a}, \mu')$. An alternative is

$$\hat{a}' = \underset{w}{\operatorname{argmin}} \ \mu'\|w\|_0 + \frac{1}{2}\|w - \hat{a}\|_2^2$$

where the L^0 'norm' $\|\cdot\|_0$ counts the number of nonzero entries. In this case, the solution is given by hard thresholding $\hat{a}$— setting all coefficients smaller in magnitude than some threshold equal to zero.

Soft thresholding is contractive and benefits from many desirable smoothing properties which make it preferable for the sparse approximation of the solution, which will be discussed below in Section 6. For a sparse approximation of the operator, the benefits of a particular choice of thresholding are less clear and therefore we consider both.

4 Theoretical Remarks The compressive spectral method, or sparse scheme, inherits many appealing properties of the underlying numerical method it approximates. In general, it is at least as stable as the original scheme and retains the order of accuracy.

4.1 Contraction and Linear Convergence The following two theorems show that the explicit and implicit numerical schemes are contractive. This result is similar to those found in [1].

THEOREM 4.1. *For the explicit scheme generating time steps by*

$$\hat{u}^{n+1} = \text{shrink}((I - dt\hat{L}_h)\hat{u}^n + dt\hat{f}, \mu),$$

if $||I - dt\hat{L}_h||_{op} \leq 1$ *then the iterations are contractive: i.e.* $||u^{n+1} - u^n||_2 \leq ||u^n - u^{n-1}||_2$.

Here $\|\cdot\|_{op}$ denotes the ℓ^2 operator norm, or largest singular value.

THEOREM 4.2. *For the implicit scheme, if* $\hat{L}_h$ *is positive semidefinite, then the iterations are contractive,* $||u^{n+1} - u^n||_2 \leq ||u^n - u^{n-1}||_2$, *for all* $dt > 0$.

The proofs of these two theorems are similar, and reside in the appendix.

The method is also convergent. In particular, for the correct scaling of μ, we have the following theorem.

THEOREM 4.3 (Linear Convergence, Explicit Scheme). *Let* S *denote a linear spectral numerical update scheme, generating time steps as*

$$\hat{u}^{n+1} = Q(\hat{u}^n, \ldots, \hat{u}^{n-k}),$$

and let S_μ *denote the spectrally sparse scheme, which generates time steps as*

$$\hat{u}_\mu^{n+1} = \text{shrink}(Q(\hat{u}_\mu^n, \ldots, \hat{u}_\mu^{n-k}), \mu).$$

Then if S *is consistent and stable (and hence converges), and if* $\mu = O(dt^{1+\delta})$ *for some* $\delta > 0$, *then the compressive scheme* S_μ *converges. If* $\mu = O(dt^p)$ *with* p *at least the order of the local truncation error of* S, *then the order of convergence of* S *is not impacted.*

For the implicit scheme, the analogous theorem is the following.

THEOREM 4.4 (Linear Convergence, Implicit Scheme). *Let* S *denote an implicit linear spectral numerical update scheme for the PDE* $u_t + Lu = f$ *on a domain* Ω *discretized with* N *grid points, generating time steps as*

$$(I + dt\hat{L}_h)\hat{u}^{n+1} = \hat{u}^n + dt\hat{f}_h,$$

and let S_μ *denote the spectrally sparse scheme, which generates time steps according to (3.2):*

$$\hat{u}^{n+1} = \underset{w}{\text{argmin}} \ \mu\|w\|_1 + \frac{1}{2}w^T(I + dt\hat{L}_h)w - w^T(\hat{u}_\mu^n + dt\hat{f}_h).$$

Then if S is consistent and stable (and hence converges), and if $\mu = O(dt^{1+\delta})$ for some $\delta > 0$, then the spectrally sparse scheme S_μ converges. If $\mu = O(dt^p)$ with p at least the order of the local truncation error of S, then the order of convergence of S in L^2 is not altered.

The proofs can be found in the Appendix.

4.2 Sparse Operator Approximation: Implicit Solver We now consider the error incurred by the sparse operator approximation proposed in Section 3.2. The continuum case is discussed in detail, and the proof for the case of discretized operators is completely analogous.

The usual discretization in Fourier space of a general, anisotropic, divergence form elliptic operator

$$Lu = -\mathrm{div}(A(x)\nabla u) + b(x)\cdot\nabla u + c(x)u$$

results in a matrix (corresponding to convolution) which is dense. However, it is still approximately sparse when the coefficients A and b are. Approximating A and b by A' and b' which are truly sparse in Fourier space yields an operator which is far more efficient to store and to work with, but incurs some error. Theorem 4.5 quantifies this error.

THEOREM 4.5. *Let u_1 and u_2 be solutions to*

$$-\mathrm{div}(A_1\nabla u_1) + b_1\cdot\nabla u_1 + c_1 u_1 = f \tag{4.1a}$$

$$-\mathrm{div}(A_2\nabla u_2) + b_2\cdot\nabla u_2 + c_2 u_2 = f. \tag{4.1b}$$

on a domain $\Omega \subset \mathbb{R}^d$ with periodic boundary conditions and the constraint

$$\int_\Omega u_1 = \int_\Omega u_2 = 0.$$

Require also that

$$w^T A_i w \geq \lambda \|w\|^2,$$
$$c_i - \frac{1}{2}\mathrm{div}(b_i) \geq 0$$

for $i = 1, 2$. Then

$$\|u_1 - u_2\|_{H^1} \leq C\lambda^{-2}\Big(d\max_{i,j}\|(\hat{A}_1)_{ij} - (\hat{A}_2)_{ij}\|_1 + C\max_i\|(\hat{b}_1)_i - (\hat{b}_2)_i\|_1 + C^2\|\hat{c}_1 - \hat{c}_2\|_1\Big)\|f\|_2$$

where $C = C(\Omega)$ is the constant from Poincare's inequality [7], and the Fourier series of the matrices and vectors A_i and b_i are taken entry-wise.

This form, in terms of $\|(\hat{A}_1)_{ij} - (\hat{A}_2)_{ij}\|_1$, $\|(\hat{b}_1)_i - (\hat{b}_2)_i\|_1$, and $\|\hat{c}_1 - \hat{c}_2\|_1$ is particularly useful because the coefficients will be approximated in Fourier space. The reader familiar with energy estimates for elliptic equations will see that the requirements of the theorem are not the most general possible, and the proof can be modified to handle other cases individually when different estimates are desired.

Proof. Subtracting the first equation of (4.1) from the second, then adding and subtracting $A_1\nabla u_2$, $b_1\nabla u_2$, and c_1u_2 gives

$$-\mathrm{div}(A_1\nabla w) - \mathrm{div}[(A_1 - A_2)\nabla u_2] + b_1\cdot\nabla w + (b_1 - b_2)\nabla u_2 + c_1 w + (c_1 - c_2)u_2 = 0,$$

and after multiplying by w and integrating by parts, we glean

$$\begin{aligned}\lambda\int_\Omega |\nabla w|^2\,dx &\le \int_\Omega \nabla w^T A_1\nabla w + \left(c_1 - \frac{1}{2}\mathrm{div}(b_1)\right)w^2\,dx\\ &\le \|A_1 - A_2\|_{op}\|\nabla u_2\|_2\|\nabla w\|_2 + \|b_1 - b_2\|_\infty\|\nabla u_2\|_2\|w\|_2 + \dots\\ &\quad \|c_1 - c_2\|_\infty\|u_2\|_2\|w\|_2.\end{aligned}$$

Using Poincare's inequality and $\|A_1 - A_2\|_{op} \le d\|A_1 - A_2\|_\infty$,

$$\lambda\int_\Omega |\nabla w|^2\,dx \le \big(d\|A_1 - A_2\|_\infty + C\|b_1 - b_2\|_\infty + C^2\|c_1 - c_2\|_\infty\big)\,\|\nabla u_2\|_2\|\nabla w\|_2$$

and thus

$$\|\nabla w\|_2 \le \lambda^{-1}\big(d\|A_1 - A_2\|_\infty + C\|b_1 - b_2\|_\infty + C^2\|c_1 - c_2\|_\infty\big)\,\|\nabla u_2\|_2. \tag{4.2}$$

Similarly multiplying the equation for u_2 by u_2, integrating by parts, and applying Poincare's inequality yields

$$\|\nabla u_2\|_2 \le C\lambda^{-1}\|f\|_2.$$

Substituting this into (4.2) and using Poincare's Inequality again, we get

$$\|u_1 - u_2\|_{H^1} \le C\lambda^{-2}\big(d\|A_1 - A_2\|_\infty + C\|b_1 - b_2\|_\infty + C^2\|c_1 - c_2\|_\infty\big)\,\|f\|_2.$$

The form stated in the theorem follows after

$$\|A\|_\infty = \max_{i,j}\|A_{ij}\|_\infty \le \max_{i,j}\|\hat{A}_{ij}\|_1$$

and the analogous inequality with b. ∎

In practice memory is not a concern due to the convolutional structure of the matrix $\hat{L}_h$ representing an elliptic operator in Fourier space, but the sparse structure of the operator dramatically reduces computation complexity (Section 7.2).

4.3 Sparse Operator Approximation: Explicit Solver The discrete analogue of Theorem 4.5 covers numerical schemes with implicit time steps, each of which require solving an elliptic problem with a sparsely approximated operator. Effectively, it allows us to estimate

$$\|Q^{-1} - P^{-1}\|_{op}$$

where P is a sparse matrix approximating that of the discretized full elliptic operator Q, and $\|\cdot\|_{op}$ refers to the L^2 matrix operator norm, or largest singular value. On the other hand, for explicit schemes, we are concerned about

$$\|Q - P\|_{op}$$

which we will consider directly.

THEOREM 4.6. *Let L be the elliptic operator defined*

$$Lv = -\mathrm{div}(a\nabla v)$$

and let Q be its Fourier discretization

$$Q\hat{u} = k\,\hat{a}_h * (k\,\hat{u})$$

where k denotes the vector of Fourier mode frequencies and a_h is the discretized domain inhomogeneity coefficient in the elliptic operator. Analogously, let

$$P\hat{u} = k\,\hat{a}'_h * (k\,\hat{u}).$$

Then

$$\|Q - P\|_{op} \leq K^2 \|\hat{a}_h - \hat{a}'_h\|_1$$

where K is the highest frequency on the grid.

In the case of a square grid $[1, \ldots, N]^d$, $K = N/2$. The result may be dismaying at first glance because it appears that the approximation error $\|\hat{a_h} - \hat{b_h}\|_1$ must be decreased faster than $O(1/N^2)$ just to remain stable. However, this type of bound is natural, since the operators' norms themselves are

$$\|P\|_{op} \approx \|Q\|_{op} = O(K^2).$$

The large operator norm is normalized by the stability condition $dt = O(dx^2)$, so one can think of these bounds in the update sense as:

$$\|(I - dtQ) - (I - dtP)\|_{op} \leq \|\hat{a}_h - \hat{a}'_h\|_1.$$

Proof. The result is a simple consequence of Young's inequality: $\|f * g\|_2 \leq \|f\|_1 \|g\|_2$. We have

$$\begin{aligned}
\|Q - P\|_{op} &= \sup_{||\hat{u}||=1} \|(Q - P)\hat{u}\|_2 \\
&= \|k(\hat{a}_h - \hat{a}'_h) * (k\hat{u})\|_2 \\
&\leq \|k(\hat{a}_h - \hat{a}'_h)\|_1 \|k\hat{u}\|_2 \\
&\leq K^2 \|\hat{a}_h - \hat{a}'_h\|_1.
\end{aligned}$$

∎

The proof clearly generalizes to hyperbolic operators of the form $Q\hat{u} = \hat{a} * (ik\,\hat{u})$ as well.

For an example, recall the forward Euler discretization of a parabolic PDE:

$$\hat{u}^{n+1} = (I - dt\hat{L}_h)\hat{u}^n,$$

over a time interval $[0, T]$. If $\|\hat{a}_h - \hat{a}'_h\|_1 = \delta$, approximating Q by P incurs an additional local truncation error of magnitude $\delta K^2 dt$ at each time step. As the grid is refined, the CFL condition requires that $K^2 dt$ stay approximately constant, so that the approximation error per step remains approximately constant.

5 The Modified Equation Prespective Using the variational principle for the explicit scheme applied to the parabolic equation yields the following first order optimality condition:

$$0 \in \hat{u}^{n+1} - (1 - dt\hat{L}_h)\hat{u}^n - dt\hat{f}_h + \mu\partial\|\hat{u}^{n+1}\|_1,$$

which is equivalent to

$$0 \in \frac{\hat{u}^{n+1} - \hat{u}^n}{dt} + \hat{L}_h\hat{u}^n - \hat{f}_h + \frac{\mu}{dt}\partial\|\hat{u}^{n+1}\|_1.$$

Taking $\mu = \delta dt$ and formally sending dt and h to zero leads to

$$\hat{u}_t + \hat{L}\hat{u} - \hat{f} \in -\delta\partial\|\hat{u}\|_1, \tag{5.1}$$

or

$$\hat{u}_t + \hat{L}\hat{u} = \hat{f} - \delta p(\hat{u}) \tag{5.2}$$

where $p(\hat{u})$ denotes the particular element of the subdifferential so that the differential inclusion (5.1) is an equality. The sparse scheme applied to hyperbolic and elliptic problems yields analogous modified equations. We consider this to be the modified equation in the sense that the numerical scheme is directly solving this problem. The subgradient contribution is a vanishing 'compression' term which may be interpreted as a force which pushes the solution u toward the nearest (in the L^1 proximal sense) union of low dimensional subspaces spanned by the Fourier basis.

Well-posedness for the modified equation is guaranteed via the theory of differential inclusions on Banach spaces (*e.g.* [1, 4]). The theorem below summarizes these results in the current context.

THEOREM 5.1 (Well-posedness). *Let $u(t)$ satisfy the differential inclusion*

$$\partial_t u(t) \in -A(u(t)) - \delta\partial||\hat{u}(t)||_{L^1}$$

with $u(0)$ in the domain of the monotone (single-valued) operator A. Then for all $\delta \geq 0$, there exists a unique solution $u(t)$ defined for all $t \geq 0$ which is the solution to

$$\partial_t u(t) = -A(u(t)) - \delta p(\hat{u}(t))$$

for some $p \in \partial||\hat{u}(t)||_{L^1}$.

Lastly, we mention that if we want to directly compare the error between the solutions of the original and modified equations, the error grows linearly in time (at worst).

THEOREM 5.2. *Let u be the solution to*

$$u_t + Lu = f$$

and let u_δ solve

$$(\hat{u}_\delta)_t + \hat{L}\hat{u}_\delta - \hat{f} \in \delta\partial\|\hat{u}_\delta\|_1.$$

Then

$$||u(t,\cdot) - u_\delta(t,\cdot)||_2 \leq 2\delta t.$$

The proof is direct and can be found in the Appendix. Similar results are easily proved for the elliptic and hyperbolic cases using only that $\|p(\hat{u})\|_\infty \leq 1$ and standard energy estimates; this approach also provides a simple alternate proof.

6 Denoising Perspective Soft thresholding also appears in early methods for signal denoising using wavelets [6]. We refer the reader to that work for full details, and list here only the analogues of its major results in the current context.

Consider the following denoising problem: we wish to recover a signal $f \in \mathbb{R}^n$ from noisy observations $d = f + w$, $||w||_1 \leq \mu$, by soft-thresholding DFT coefficients by μ. This approach enjoys the following properties:

- (Smoothing) The recovered signal f_μ satisfies $||f_\mu||_{H^k} \leq ||f||_{H^k}$ for any Sobolev norm $\|\cdot\|_{H^k}$. In particular, $|\hat{f_\mu}(k)| \leq |\hat{f}(k)|$ for all frequencies k.
- (Near optimality) f_μ is near-minimax:

$$\sup_{||f||_{H^k} \leq C_1} \sup_{||w||_1 \leq \mu} ||f_\mu - f||_{l_n^2}^2 \leq 4 \inf_{\tilde{f}} \sup_{||f||_{H^k} \leq C_1} \sup_{||w||_1 \leq \mu} ||\tilde{f}(d) - f||_{l_n^2}^2$$

  where $\tilde{f}(d)$ is any other estimator of f.

The smoothing property guarantees that the recovered signal is 'noise-free'; the near optimality property guarantees that for worst-case signals of bounded Sobolev norm and noise of bounded ℓ^1 norm, the result recovered by soft thresholding is nearly the 'optimal' (see [6]).

Next, consider the solution u^ϵ to the standard parabolic multiscale problem

$$\frac{\partial u^\epsilon}{\partial t} - \frac{\partial}{\partial x}\left(a(x, x/\epsilon)\frac{\partial u^\epsilon}{\partial x}\right) = 0 \quad \text{on } [0, 2\pi] \text{ periodic}, \quad u^\epsilon(x, 0) = u_0^\epsilon(x).$$

The theory of asymptotic homogenization (*e.g.* [11]) can be used to show that at time point t^n, the exact solution u^ϵ satisfies

$$u^\epsilon(x, t^n) = u_0(x, t^n) + \epsilon u_1(x, x/\epsilon, t^n) + \epsilon^2 R(x, t^n)$$

with $|\hat{R}(x, t)| \leq C$. This expansion is valid as long as we assume that the equation is taken on a periodic domain and $a(x)$ is as smooth as we like. For a numerical solution, the asymptotic expansion can be easily modified to include truncation error τ^{n+1} as follows: if we let v^{n+1} denote the numerical solution at time t^{n+1}, then

$$v^{n+1} = u_0(x, t^{n+1}) + \epsilon u_1(x, x/\epsilon, t^{n+1}) + \epsilon^2 R(x, t^{n+1}) - \tau^{n+1}.$$

This form allows us to draw a connection between the denoising and homogenization problems: for an appropriate threshold choice μ, the compressive spectral method denoises v^{n+1} as

$$v^{n+1} = \underbrace{u_0(x, t^{n+1}) + \epsilon u_1(x, x/\epsilon, t^{n+1})}_{\text{signal}} + \underbrace{\epsilon^2 R(x, t^{n+1}) - \tau^{n+1}}_{\text{noise}}$$

and attempts to recover the first terms in the asymptotic expansion. These interpretation is valid between any two time steps, but may not hold globally.

7 Efficient Implementation In this section, we describe important details pertaining to the numerical method and algorithm considerations. Using a concrete example, we show that a favorable complexity can be achieved.

7.1 The Proximal-Galerkin Algorithm The implicit scheme described above requires fast minimization of the energy (3.2), and differs from many cases where L^1 regularization is added because the problem, *e.g.* compressed sensing [2], TV minimization [12], or basis pursuit [3], is ill-posed without it. For the multiscale

PDE problem, this is not the case since an appropriately discretized version of (3.2) will be well-posed and can be solved by inverting a linear system

$$Q\hat{u} = \hat{f}$$

where Q is a positive-definite (and even sparse, in physical rather than Fourier space) matrix. If the elliptic operator is discretized appropriately, fast and extensively studied preconditioned conjugate gradient solvers are available. So, to be competitive, the compressive implicit scheme must leverage sparsity of the solution $\hat{u}$ to perform the (approximate) linear inversion $Q\hat{u} = \hat{f}$ quickly. For this purpose, we propose the hybrid proximal gradient descent and Galerkin approximation algorithm described below, which is related to the procedure described in [5].

First, let D be the diagonal part of Q. Since Q is the matrix corresponding to a Fourier-space discretized elliptic operator, D is the matrix corresponding to a multiple the Fourier-space discretized Laplacian. We take $n \sim 10$, $\mu > 0$, $\omega > 0$, and initialize the solution to be zero (*i.e.* $\hat{u} = 0$).

The Proximal-Galerkin Algorithm

for $j = 1{:}n$ **do**
  $\hat{u} = \text{shrink}\big(\hat{u} + \omega D^{-1}(\hat{f} - Q\hat{u}), \mu\big)$;
end for
set $I = \text{supp}(\hat{u})$;
set $\hat{u} = \underset{w:\ \text{supp}(w)\subseteq I}{\text{argmin}} \frac{1}{2}\|Qw - \hat{f}\|_2^2$;
Return $\hat{u}$.

The algorithm begins with a few iterations of the proximal gradient method applied to the energy

$$E(w) = \mu\|w'\|_1 + \frac{1}{2}w'^T Q' w' - w'^T \hat{f}'$$

where

$$\begin{aligned} Q' &= D^{-1/2} Q D^{-1/2}, \\ \hat{f}' &= D^{-1/2}\hat{f}, \\ w' &= D^{1/2} w. \end{aligned}$$

This is a simple Jacobi preconditioning of the analogous energy with Q, w, and $\hat{f}$. Rather than iterating proximal gradient to convergence, which would be too slow, the algorithm stops after just a few iterations with rough approximation. The support of that solution is used to identify the Fourier modes with largest magnitude coefficients, and then a Galerkin approximation is computed over those modes. Due to sparsity in the Fourier domain, the linear solve associated with the Galerkin part is small and inexpensive– computational complexity depends on the grid mesh size only through the sparsity of the solution.

7.2 Algorithm Complexity The pseudospectral approach of computing the convolution

$$k\,\hat{a} * (k\,\hat{u})$$

uses an FFT, and for an N-gridpoint problem this reduces the computational complexity per iteration from $O(N^2)$ to $O(N \log N)$. We now consider the computational complexity of the sparse spectral method, which must be comparable to $O(N \log N)$ to be practical.

Suppose that the sparsely approximated operator is defined $P\hat{u} = k\,\hat{a}' * (k\,\hat{u})$, where the sparsity (number of nonzeros) of $\hat{a}'$ is m, and that the sparsity of $\hat{u}$ is r. By treating the $\hat{a}' * \hat{u}$ sparse convolution as a summation of sparse vectors, it can be accomplished with complexity

$$O(mr \min(\log r, \log m)), \tag{7.1}$$

free of any dependence on the full problem size N, by storing the sparse vectors $\hat{a}'$ and $\hat{u}$ as sorted linked lists and computing the sum as a merge operation, with a priority queue. For the modest one-time cost of initializing a length N array, the complexity can be decreased to

$$O(mr) \tag{7.2}$$

by leaving the sparse vectors unsorted. We iterate over the mr nonzero coefficients which must be added, and use an auxiliary array keep track of the partial result. When a new coefficient of the partial result becomes nonzero, it is placed in a growing list of indices. After we have visited each of the mr coefficients to be added, we iterate over the list of nonzero indices, perform the shrink operation on the corresponding auxiliary array entry holding the partial result, and copy the outcome into a list which holds the final result. Along the way, we 'zero out' the entry of the partial result array, never incurring another $O(N)$ cost.

Finally, if the problem is elliptic or requires implicit time steps and the Proximal-Galerkin algorithm is used, the complexity includes a term

$$O(r^3),$$

the cost of the Galerkin linear solve over the support found with proximal gradient.

Both (7.1) and (7.2) are preferable to the $O(N \log N)$ cost of the pseudospectral method for very sparse problems and in the homogenization limit discussed next in Section 7.3. For the numerical examples considered in this limit, m and r stay approximately constant, leading to computation time which does not increase as the grid is refined.

One key to the effective application of the sparse spectral method is proper discretization. For a typical homogenization problem, we are interested in the solution of an equation such as

$$-\mathrm{div}(a(x/\epsilon)\nabla u) = f$$

for ϵ close to zero, and we might choose the inhomogeneity coefficient

$$a(x) = 1 + \frac{1}{2}\sin \pi x.$$

This choice is ideally sparse in the Fourier domain, with only three nonzero entires regardless of N, using the standard uniform grid. If $\epsilon = 1/1000$, then $\hat{a}$ still has only three nonzeros. However, choosing $\epsilon = \frac{1}{707\sqrt{2}}$ results in $\hat{a}$ being completely dense. These two choices of ϵ differ by less than 10^{-6}, and the first leverages extreme sparsity in the problem while the second does not. This example shows that it is prudent to assume a certain relationship between the grid spacing and ϵ, considered next.

7.3 Homogenization Limit For homogenization problems in particular, where one is interested in the limit $\epsilon \to 0$, we can keep $N\epsilon$ fixed as the grid is refined. Empirically, we have observed that this keeps the sparsity of the operator and of the solution approximately constant. For a simple case of this $N\epsilon = c$ (c a constant) limit, the following theorem guarantees the sparsity of the operator remains fixed along a subsequence.

THEOREM 7.1. *Let L_ϵ be the elliptic operator defined*

$$L_\epsilon v = -\mathrm{div}(a(x/\epsilon)\nabla v),$$

and let

$$Q_{\epsilon,N} u = k\, \hat{a_N} * (k\, \hat{u})$$

be its Fourier discretization on an $N-$point discretization of $[0, 2\pi)$. Then $Q_{\epsilon,N}$ and $Q_{\epsilon/2,2N}$ are equally sparse: that is,

$$\#\{k : |\hat{a_{2N}}(k)| \geq \lambda\} = \#\{k : |\hat{a_N}(k)| \geq \lambda/2\} \tag{7.3}$$

for all $\lambda > 0$.

See the appendix for a proof. Note that the theorem assumes the standard definition of the DFT on N grid points,

$$\mathcal{F}_N[a(x)](k) = \sum_{j=0}^{N-1} a\left(\frac{2\pi j}{N}\right) e^{-2\pi i j k/N},$$

which is not unitary. This accounts for the appearance of $\lambda/2$ rather than λ on the right hand side of (7.3). This factor cancels out in the end because with this definition of the DFT, the ℓ^1 norm in Theorems 4.5 and 4.6 should be scaled by $1/N$.

The complexities (7.1) and (7.2) become very favorable in the $N\epsilon = c$ limit, where m and r remain nearly constant or grow approximately logarithmically with N as the grid is refined. In each case we observed, the overall algorithm complexity is linear or sub-linear in N.

8 Numerical Examples In [13], the authors demonstrated the effective application of the compressive spectral method to a variety of problems. Here, we expand on those results and give examples of the additions to the method proposed in this paper: the implicit scheme and sparse operator approximation.

8.1 Transport Equation, 1D The PDE considered is the traveling wave equation:

$$\begin{gathered} u_t + a(x)u_x = 0, \\ x \in [0, 2\pi] \text{ periodic}, \\ u(x, 0) = \sin(x) \end{gathered}$$

with oscillatory coefficient

$$a(x) = \frac{1}{8} \exp\left(\frac{0.6 + 0.2\cos x}{1 + 0.7 \sin 128x}\right).$$

The update is given by leap frog time discretization:

$$\hat{u}^{n+1} = \hat{u}^{n-1} - 2dt\hat{a} * (ik\, \hat{u}^n).$$

We choose the above form for a throughout this section, because it is less sparse than simple trigonometric functions.

The grid sizes considered are $N = 2^{10}, \ldots, 2^{14}$ and the values of other parameters are $dt = 6.25\times10^{-6}$, $\|\hat{a} - \hat{a}'\|_1 = 10^{-2}$, $\mu = 1.2\times10^{-5}$, and the simulation is run to a final time $t = 0.5$.

Figure 8.1 shows the full spectral and compressive spectral (sparse operator/sparse solution) solutions on coarse and fine scales. The compressive scheme correctly captures the largest Fourier coefficients of the solution, discarding all but 3.7%, and the operator approximation discards all but 2.6%. The "true" solution was computed on a fine grid with finite difference methods.

Figure 8.2 shows the L^2 error and sparsity of the compressive spectral approximations as the grid is refined with dt held constant. Error is computed as the L^2 distance to the full spectral solution. The error of the sparse operator/sparse solution scheme is dominated by the sparse approximation of the solution; spurious modes in the leap frog scheme make a sparse approximation of it difficult. Over the range of grids considered, sparsity of the operator eventually becomes constant while sparsity of the solution grows about linearly. The complexity of the compressive spectral method is thus linear in N over the grid sizes considered.

Figure 8.3 considers the same problem but with a resonant forcing term

$$f(x) = e^{\sin(x/128)^2}$$

with $N = 2048$ and all other parameters the same as the non-forced problem. The solution has 11.3% nonzero Fourier coefficients, with $\|u_{\text{full}} - u_{\text{sparse}}\|_2 = 2.5\times10^{-3}$. The resonant forcing causes sharp and irregular oscillations at the fine scale, which make the problem less sparse, but the compressive scheme still captures the correct behavior.

8.2 Elliptic Problem, 1D The PDE considered is the elliptic problem:

$$\begin{gathered} -(a(x)u_x)_x = \sin 2x, \\ x \in [0, 2\pi] \text{ periodic}, \\ \int u\,dx = 0 \end{gathered}$$

with

$$a(x) = \exp\left(\frac{0.6 + 0.2\cos x}{1 + 0.7\sin x/\epsilon}\right)$$

such that $N\epsilon = 8$, and usual spectral operator discretization:

$$\hat{L}_h\hat{u} = k\,\hat{a}_h * (k\,\hat{u}) = \hat{f}.$$

This time we consider the homogenization limit, keeping $N\epsilon = 8$ with $\epsilon = \frac{1}{64}, \frac{1}{128}, \ldots, \frac{1}{1024}$, and set $\|\hat{a} - \hat{a}'\|_1 = 1\times10^{-4}$. Parameter values for the Proximal-Galerkin algorithm are $n = 10$, $\mu = 5\times10^{-8}$, and $\omega = 5\times10^{-3}$.

Figure 8.4 shows the full spectral and compressive spectral solutions on coarse and fine scales. Both the sparse solution and operator approximation keep 8.5% of the coefficients. Note that the full result and sparse operator/sparse solution result lie almost on top of each other, even at the resolution of the fine scale.

Figure 8.5 shows error (L^2 distance to the full spectral solution) and sparsity under refinement. "Sparse operator" refers to the solution obtained with the sparsely

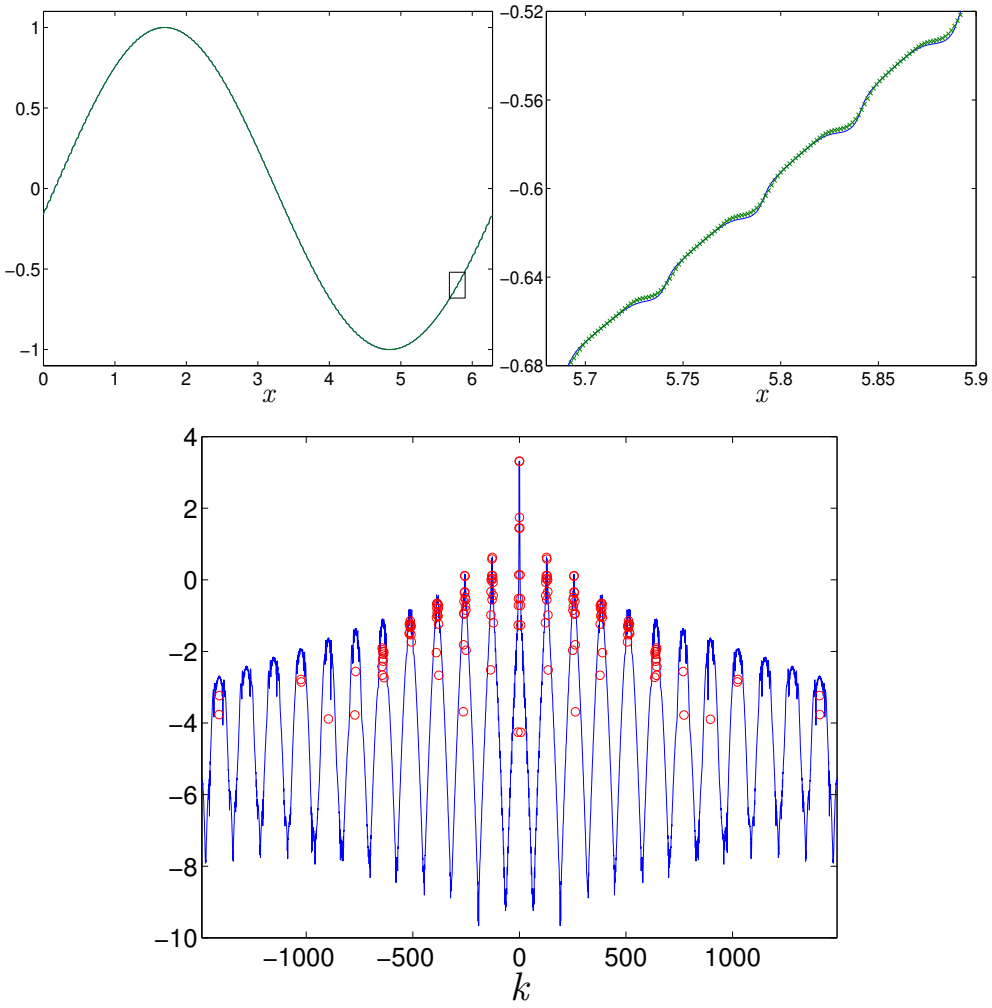

Fig. 8.1: **Left:** True (blue) and sparse operator/sparse solution (green) solutions in physical space. The two curves lie almost on top of each other. **Right:** Zoomed in true (blue) and sparse (green '×') solutions. **Bottom:** True (blue) and sparse (red '∘') solutions in Fourier space. $N = 4096$, operator nonzeros = 107, solution nonzeros = 153.

approximated operator, using either a high accuracy conjugate gradient solve or the Proximal-Galerkin algorithm. "Sparse solution" refers to the use of the Proximal-Galerkin algorithm, with either the full or sparse operator.

Approximation error does not increase while both solution and operator sparsity remain approximately constant, leading to computation time approximately independent of N. With $N = 2^{13}$, the sparse approximation maintains six digits of accuracy with only 1.1% of the coefficients of both the operator and the solution.

Figure 8.6 illustrates that for a fixed number of nonzero coefficients, the sparse operator approximation incurs smaller error than the solution approximation.

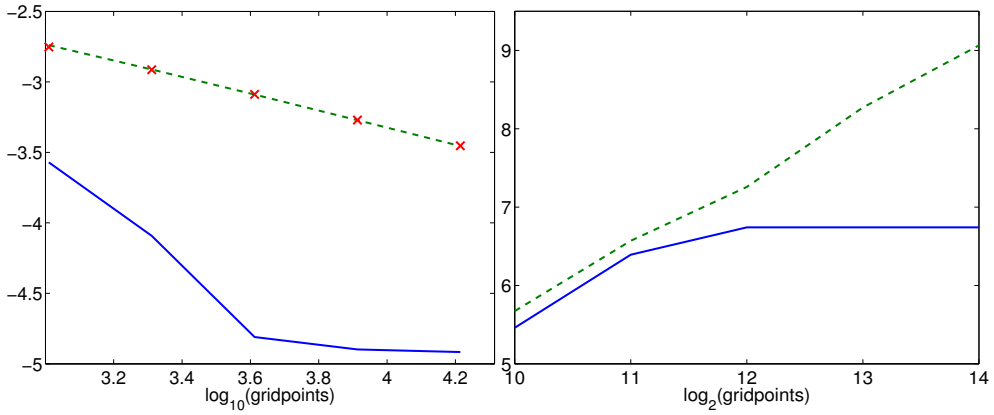

Fig. 8.2: **Left:** Sparse operator/full solution (blue), full operator/sparse solution (green, dashed), and sparse operator/sparse solution (red $\times$) L^2 distance to the full spectral solution as the grid is refined. The y axis has a $\log_{10}$ scale. **Right:** Number of nonzero Fourier coefficients of the operator (blue) and solution (green, dashed) as the grid is refined. The y axis has a $\log_2$ scale.

8.3 Parabolic Problem, 1D The PDE we consider here is the parabolic equation:

$$\begin{aligned} u_t - (a(x)u_x)_x &= 0, \\ x &\in [0, 2\pi] \text{ periodic}, \\ u(x,0) &= 1 + \cos(x - \pi) \end{aligned}$$

with

$$a(x) = \exp\left(\frac{0.6 + 0.2\cos x}{1 + 0.7\sin x/\epsilon}\right)$$

We again consider the $N\epsilon = 8$ limit, $\epsilon = \frac{1}{64}, \frac{1}{128}, \ldots, \frac{1}{1024}$, and set $\|\hat{a} - \hat{a}'\|_1 = 1\times10^{-2}$ and $dt = 1\times10^{-2}$ for all N. Parameter values for the Proximal-Galerkin algorithm are $n = 10$, μ ranges from 5×10^{-6} to 6.4×10^{-6}, and $\omega = 1\times10^{-2}$.

Figure 8.7 compares the solutions on coarse and fine scales. The sparse solution retains 3.2% of the coefficients and the operator is also approximated with 3.2%. Figure 8.8 shows error and sparsity under refinement. Approximation error decreases while sparsity of both operator and solution stay constant. The overall complexity is thus constant in N over the range of grid sizes considered. For this problem, sparse approximation of the operator incurs most of the error.

8.4 Elliptic Problem, 2D We consider the elliptic problem

$$\begin{aligned} -\text{div}(a(x)\nabla u) &= 10\sin x \sin y, \\ x, y &\in [0, 2\pi] \text{ periodic}, \\ \int u\, dxdy &= 0 \end{aligned}$$

with

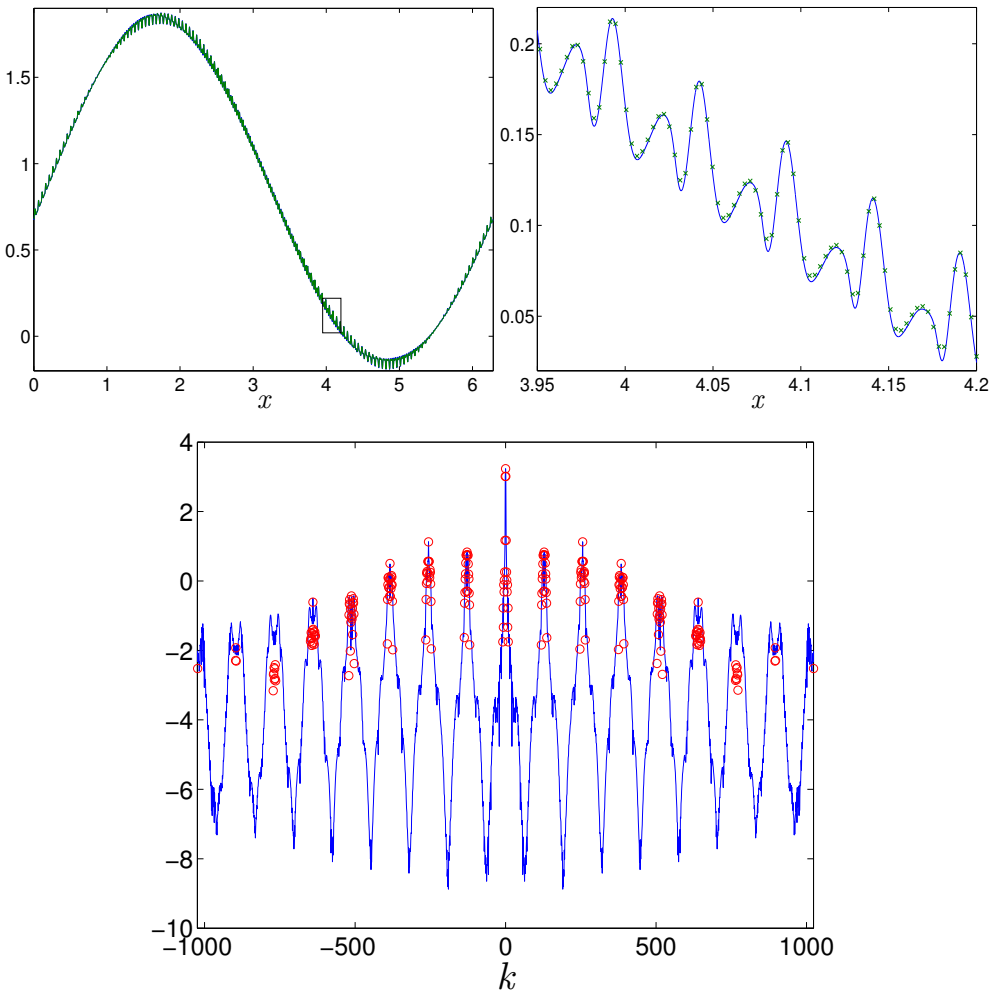

Fig. 8.3: **Left:** True (blue) and sparse operator/sparse solution (green) solutions with resonant forcing term in physical space. **Right:** Zoomed in true (blue) and sparse (green '×') solutions. **Bottom:** True (blue) and sparse (red '∘') solutions in Fourier space. $N = 2048$, operator nonzeros $= 86$, solution nonzeros $= 231$.

$$a(x,y) = \exp\left(\frac{0.6 + 0.2\cos x}{1 + 0.7\sin x/\epsilon} + \frac{0.6 + 0.2\cos y}{1 + 0.7\sin y/\epsilon}\right)$$

on an $N \times N$ grid such that $N\epsilon = 8$, with $\epsilon = \frac{1}{16}, \frac{1}{32}, \ldots, \frac{1}{256}$ and $\|\hat{a} - \hat{a}'\|_1 = 1$. Parameter values for the Proximal-Galerkin algorithm are $n = 20$, μ between $4{\times}10^{-4}$ and $32{\times}10^{-4}$, and $\omega = 2{\times}10^{-2}$.

Because the full and spectral solutions are very close to each other in physical space and an overlaid comparison of surfaces is difficult, Figure 8.9 shows the solutions on a log scale in Fourier space. Of the 2^{20} coefficients in the full solution, the sparse solution and operator retain just 0.2% while maintaining four digits of accuracy. Figure 8.10 shows that approximation error decreases slightly with constant sparsity and computation time. For some grid sizes, the sparse operator/sparse solution scheme

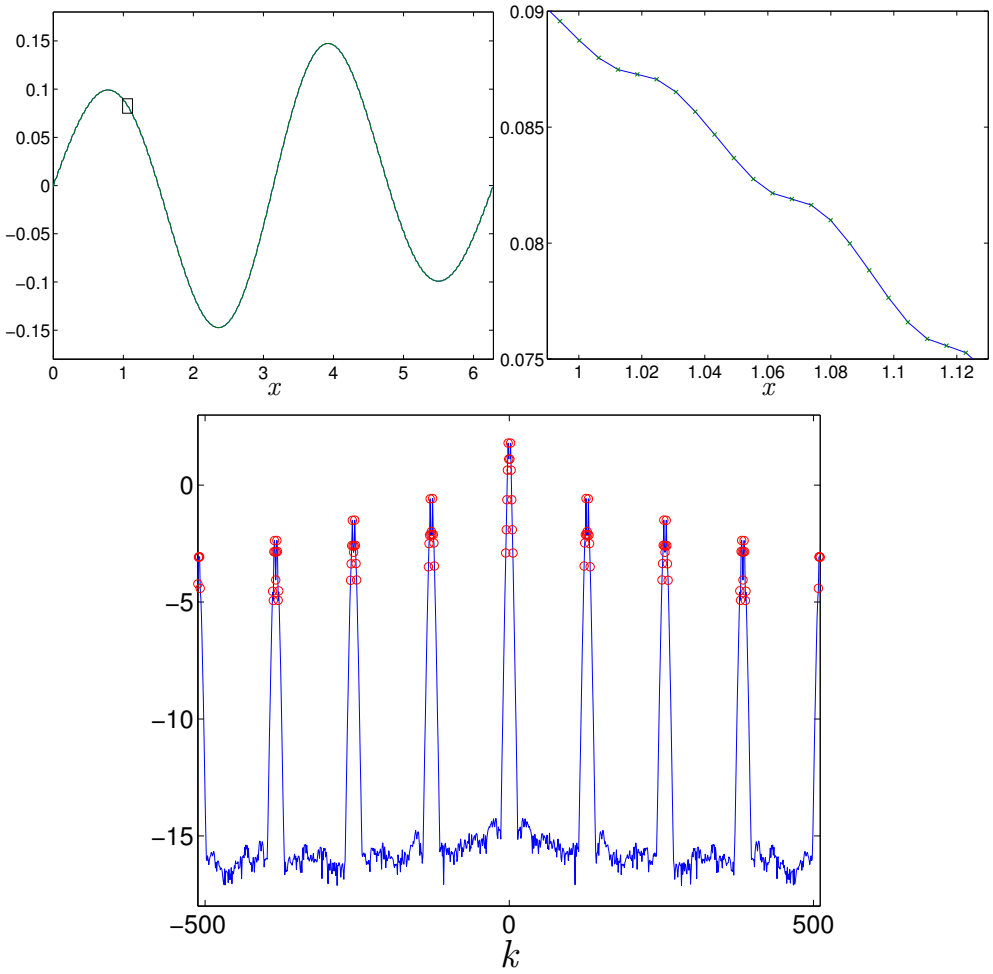

Fig. 8.4: **Left:** True (blue) and sparse operator/sparse solution (green) solutions in physical space. The small rectangle shows the axis limits of the zoomed in plot to the right. **Right:** Zoomed in true (blue) and sparse (green '×') solutions. **Bottom:** True (blue) and sparse (red '∘') solutions in Fourier space. $N = 1024$, operator nonzeros = 86, solution nonzeros = 87.

actually attains a lower error than the sparse operator/full solution scheme, evidence of the denoising effect discussed in Section 6.

To compare the Fourier coefficients of the full and sparse solutions more accurately, the left panel of Figure 8.11 shows the magnitude of the 4500 largest Fourier coefficients of the true solution sorted in descending order. The magnitude of the corresponding sparse solution Fourier coefficients is also shown, with an upward bias to account for all the wave numbers not present. The right panel shows the fraction of full solution wave numbers which are captured by the sparse scheme. The compressive scheme correctly identifies all 500 of the largest modes in the full solution, and about 68% of the full solution's largest 1800 modes.

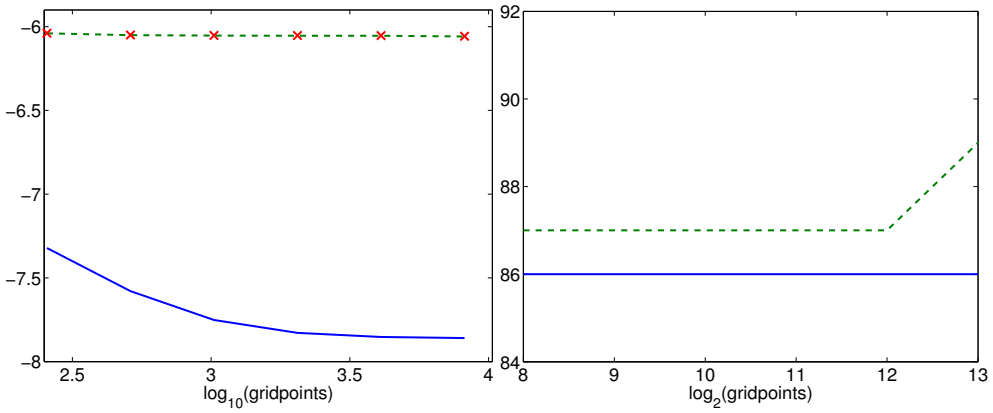

Fig. 8.5: **Left:** Sparse operator/full solution (blue), full operator/sparse solution (green, dashed), and sparse operator/sparse solution (red $\times$) error under the homogenization limit. The y axis has a $\log_{10}$ scale. **Right:** Number of nonzero Fourier coefficients of the operator (blue) and solution (green, dashed) as the grid is refined.

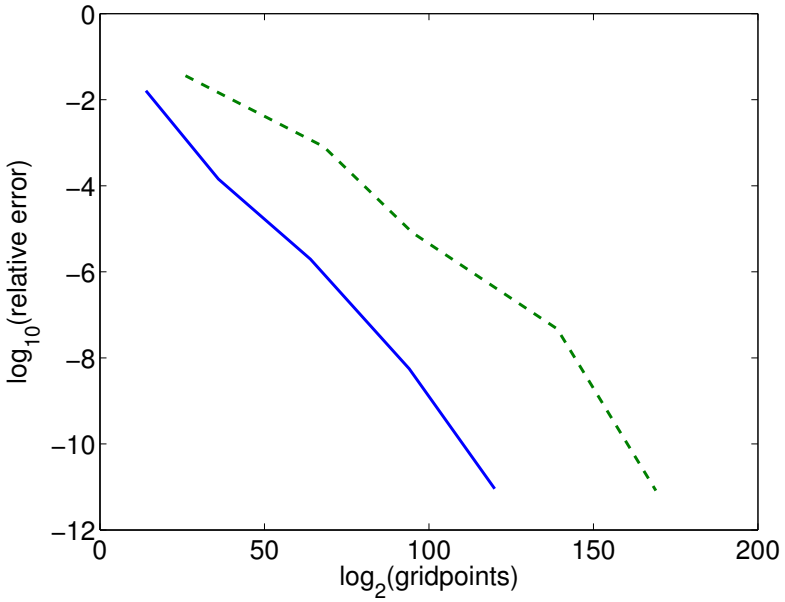

Fig. 8.6: Pareto curves showing the tradeoff between approximation error and sparsity of the operator (blue) and solution (green, dashed).

9 Conclusion In this paper, we have proposed a sparse operator approximation and an efficient method for extending the work of [13] to implicit solvers (Section 3). We have proven the convergence of the original compressive spectral scheme [13] and the new variants, including a modified equation which shows the effect of soft thresholding is equivalent to including an L^1 subgradient term in the PDE. Also,

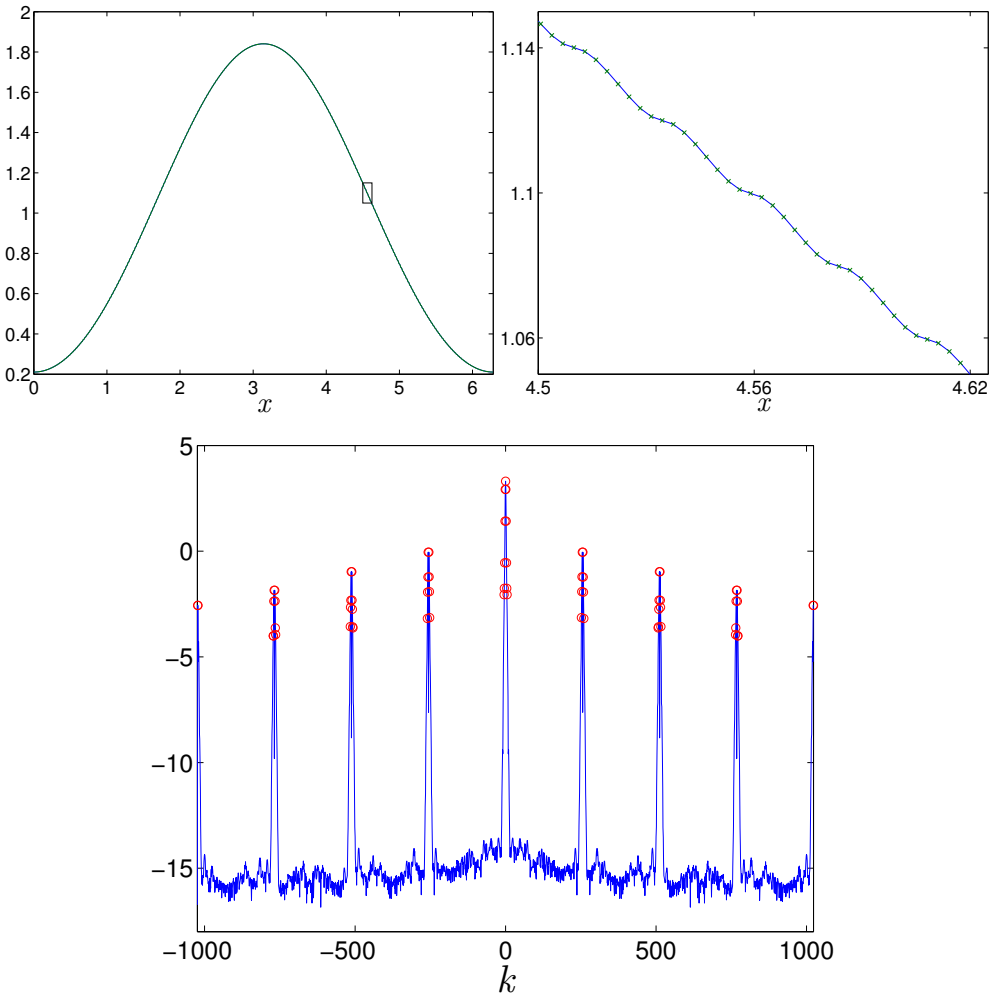

Fig. 8.7: **Left:** True (blue) and sparse operator/sparse solution (green) solutions in physical space. **Right:** Zoomed in true (blue) and sparse (green '×') solutions. **Bottom:** True (blue) and sparse (red '○') solutions in Fourier space. $N = 2048$, operator nonzeros $= 64$, solution nonzeros $= 65$.

we connect the homogenization problem with that of signal denoising via wavelet thresholding. For PDE with sparse initial data or forcing terms, the new methods are asymptotically preferable to the pseudospectral approach. The methodology presented here could be translated to other psuedospectral methods which employ alternative bases. Computationally, this amounts to replacing the Fast Fourier transforms in the psuedo-codes above with the appropriate transformation. This could be useful in cases where the solutions are sparse against another known basis.

Acknowledgments The authors would like to thank Will Feldman, Inwon Kim, Chris Anderson, and Russel Caflisch for insightful discussions regarding the above. A. Mackey was funded by UC Lab grant 12-LR-236660, and in part by the National

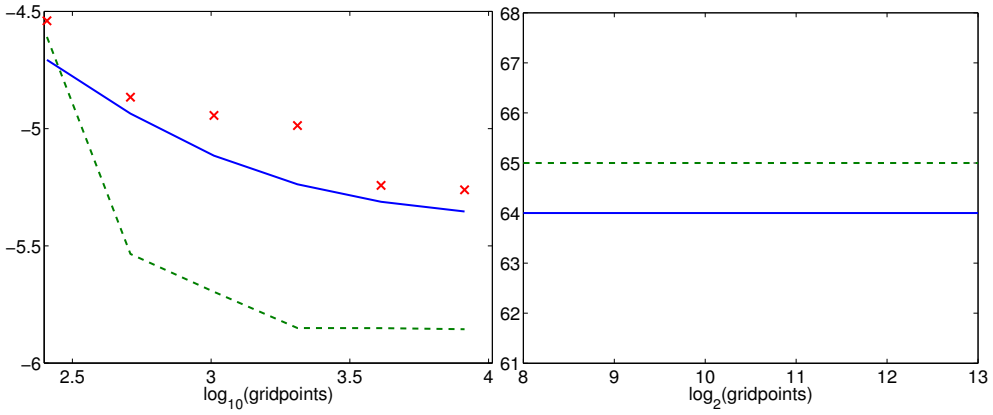

Fig. 8.8: **Left:** Approximation error of the sparse operator/full solution (blue), full operator/sparse solution (green, dashed), and sparse operator/sparse solution (red $\times$) error under the homogenization limit. The y axis has a $\log_{10}$ scale. **Right:** Number of nonzero Fourier coefficients of the operator (blue) and solution (green, dashed) are constant as the grid is refined.

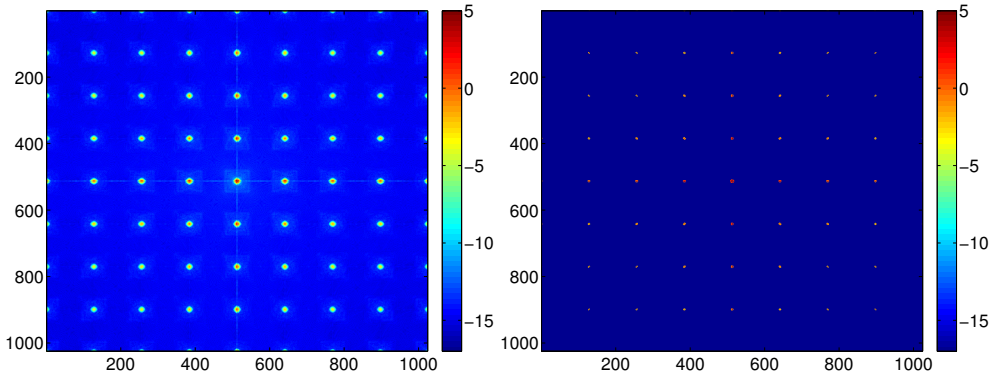

Fig. 8.9: Full (left) and sparse (right) solutions on a log scale in Fourier space. Note that the great majority of coefficients in the sparse solution are exactly zero. $N = 1024$, $\epsilon = \frac{1}{128}$, operator nonzeros $= 1972$, solution nonzeros $= 1874$.

Science Foundation through DMS 0907931. H. Schaeffer was supported by NSF DMS 1303892 and University of California President's Postdoctoral Fellowship Program. S. Osher was supported by ONR Grant N00014-11-1-719.

Appendix.

Before giving the proofs of the theorems from Section 4.1, we recall the definition of Bregman Distance (also known as Bregman Divergence).

DEFINITION 9.1. *Let J be a convex function and u, v be points in the domain of J. Also let p be an element of the subdifferential of J, i.e. $p \in \partial J(v)$. We define the*

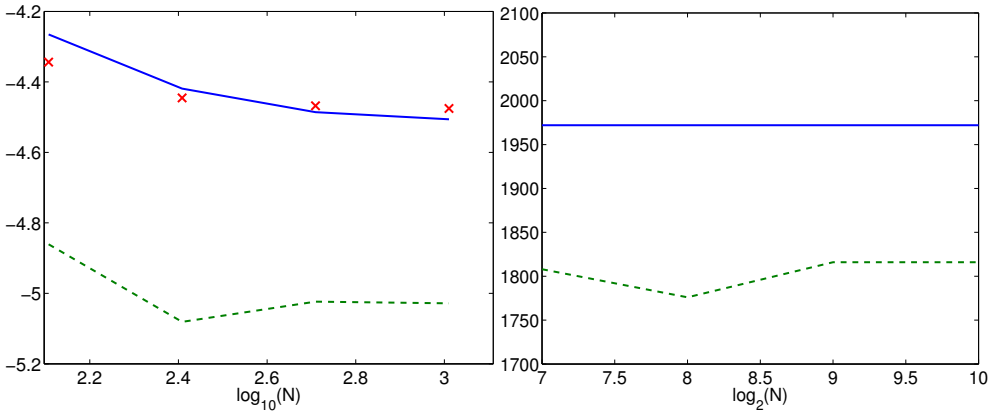

Fig. 8.10: **Left:** Approximation error of the sparse operator/full solution (blue), full operator/sparse solution (green, dashed), and sparse operator/sparse solution (red $\times$) error under the homogenization limit. The y axis has a $\log_{10}$ scale. **Right:** Number of nonzero Fourier coefficients of the operator (blue) and solution (green, dashed) are constant as the grid is refined.

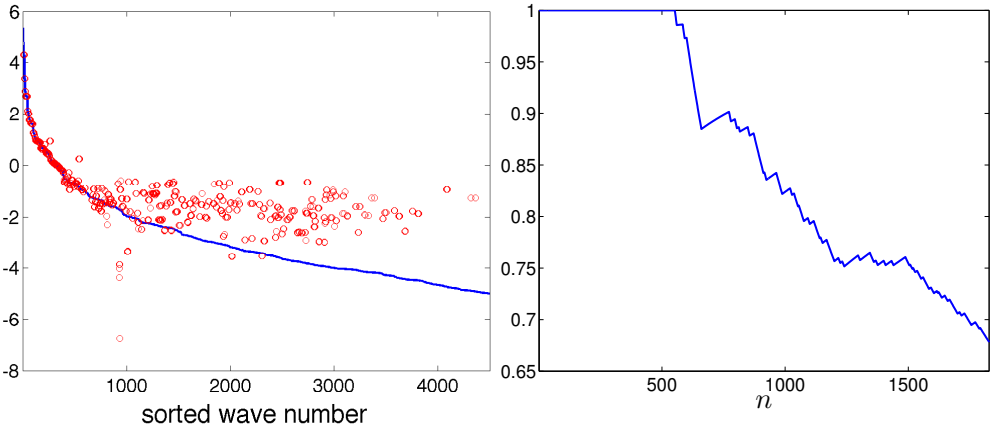

Fig. 8.11: **Left:** energy spectrum decay of the full and sparse solutions. The plot shows just the largest 4500 coefficients of the full solution, the support of which contains all coefficients of the sparse solution. **Right:** fraction of sparse modes appearing among the largest n true modes, as a function of n.

Bregman Distance *between* u *and* v *as*

$$D_J^p(u,v) = J(u) - J(v) - \langle p, u - v\rangle.$$

In general, the Bregman Distance is not symmetric and does not obey the triangle inequality, so it is not a distance in the typical sense.

In what follows, we will also use basic facts regarding monotone operators.

DEFINITION 9.2. *Let* A *be a multi-valued map from* V *into itself. We call* A monotone *if and only if for any* $u, v \in Dom(A)$ *and any values* Au *and* Av *might take*

on,

$$\langle u - v, Au - Av \rangle \geq 0.$$

If $A = \partial F$ is the subdifferential of a convex function, then it is monotone.

We can now give the proof of theorem 4.1, in which we omit hats for notational clarity.

Proof. Consider the iterations for arbitrary points u^n and v^n:

$$\mu p(u^{n+1}) + \frac{u^{n+1} - u^n}{dt} = -\hat{L}_h u^n + f$$
$$\mu p(v^{n+1}) + \frac{v^{n+1} - v^n}{dt} = -\hat{L}_h v^n + f.$$

By taking the difference between these two equations we arrive at

$$\mu(p(u^{n+1}) - p(v^{n+1})) + \frac{1}{dt}(u^{n+1} - v^{n+1}) - \frac{1}{dt}(u^n - v^n) = -\hat{L}_h(u^n - v^n)$$

and taking the inner product of this equation with $u^{n+1} - v^{n+1}$ yields

$$\mu \left\langle p(u^{n+1}) - p(v^{n+1}), u^{n+1} - v^{n+1} \right\rangle + \frac{1}{dt} \left\langle u^{n+1} - v^{n+1}, u^{n+1} - v^{n+1} \right\rangle -$$
$$\frac{1}{dt} \left\langle u^n - v^n, u^{n+1} - v^{n+1} \right\rangle = \left\langle -\hat{L}_h(u^n - v^n), u^{n+1} - v^{n+1} \right\rangle.$$

Rearranging terms and taking upper bounds we get the following:

$$\begin{aligned}
&\mu dt \left\langle p(u^{n+1}) - p(v^{n+1}), u^{n+1} - v^{n+1} \right\rangle + ||u^{n+1} - v^{n+1}||^2 \\
&\quad = \left\langle u^n - v^n, u^{n+1} - v^{n+1} \right\rangle + \left\langle -dt\hat{L}_h(u^n - v^n), u^{n+1} - v^{n+1} \right\rangle \\
&\quad = \left\langle (I - dt\hat{L}_h)(u^n - v^n), u^{n+1} - v^{n+1} \right\rangle \\
&\quad \leq ||(I - dt\hat{L}_h)(u^n - v^n)||||u^{n+1} - v^{n+1}|| \\
&\quad \leq ||(I - dt\hat{L}_h)||_{op}||u^n - v^n||||u^{n+1} - v^{n+1}||.
\end{aligned}$$

Note that $\mu dt \left\langle p(u^{n+1}) - p(v^{n+1}), u^{n+1} - v^{n+1} \right\rangle$ is non-negative by monotonicity of the subgradient of a convex function. We show this here by using the nonnegativity of Bregman distance:

$$\begin{aligned}
0 &\leq D_F^p(u^{n+1}, v^{n+1}) + D_F^p(v^{n+1}, u^{n+1}) \\
&= F(u^{n+1}) - F(v^{n+1}) - \left\langle p(v^{n+1}), u^{n+1} - v^{n+1} \right\rangle + F(v^{n+1}) - F(u^{n+1}) - \left\langle p(u^{n+1}), v^{n+1} - u^{n+1} \right\rangle \\
&= \left\langle p(u^{n+1}) - p(v^{n+1}), u^{n+1} - v^{n+1} \right\rangle.
\end{aligned}$$

Combining the positivity of the subgradient terms with equation above provides us with the following bound (assuming $||(I - dt\hat{L}_h)||_{op} \leq 1$):

$$||u^{n+1} - v^{n+1}|| \leq ||(I - dt\hat{L}_h)||_{op}||u^n - v^n|| \leq ||u^n - v^n||$$

as desired. ∎

Proof of theorem 4.2.

Proof. Considering the optimality condition for the energy (3.2) defining the implicit scheme, we see that the iterations for u^n and v^n can be written

$$\mu p(u^{n+1}) + \frac{u^{n+1} - u^n}{dt} = -\hat{L}_h u^{n+1} + f$$
$$\mu p(v^{n+1}) + \frac{v^{n+1} - v^n}{dt} = -\hat{L}_h v^{n+1} + f.$$

(If the operator $\hat{L}_h$ being considered in (3.2) is not positive semidefinite, then use (3.1) instead.) By taking the difference between these two equations we arrive at

$$\mu(p(u^{n+1}) - p(v^{n+1})) + \frac{1}{dt}(u^{n+1} - v^{n+1}) - \frac{1}{dt}(u^n - v^n) = -\hat{L}_h(u^{n+1} - v^{n+1}).$$

Next, taking the inner product of this equation with $u^{n+1} - v^{n+1}$ yields

$$\mu \left\langle p(u^{n+1}) - p(v^{n+1}), u^{n+1} - v^{n+1} \right\rangle + \frac{1}{dt} \left\langle u^{n+1} - v^{n+1}, u^{n+1} - v^{n+1} \right\rangle -$$
$$\frac{1}{dt} \left\langle u^n - v^n, u^{n+1} - v^{n+1} \right\rangle = \left\langle -\hat{L}_h(u^{n+1} - v^{n+1}), u^{n+1} - v^{n+1}. \right\rangle$$

Re-arranging terms and taking upper bounds we get the following:

$$\mu dt \left\langle p(u^{n+1}) - p(v^{n+1}), u^{n+1} - v^{n+1} \right\rangle + \|u^{n+1} - v^{n+1}\|^2$$
$$= \left\langle u^n - v^n, u^{n+1} - v^{n+1} \right\rangle + \left\langle -dt\hat{L}_h(u^{n+1} - v^{n+1}), u^{n+1} - v^{n+1} \right\rangle.$$

As in the explicit timestep case, $\left\langle p(u^{n+1}) - p(v^{n+1}), u^{n+1} - v^{n+1} \right\rangle \geq 0$ and so

$$||u^{n+1} - v^{n+1}||^2 \leq \left\langle u^n - v^n, u^{n+1} - v^{n+1} \right\rangle + \left\langle -dt\hat{L}_h(u^{n+1} - v^{n+1}), u^{n+1} - v^{n+1} \right\rangle.$$

If $\hat{L}_h$ is positive semidefinite then we have

$$||u^{n+1} - v^{n+1}||^2 \leq \left\langle u^n - v^n, u^{n+1} - v^{n+1} \right\rangle \leq ||u^n - v^n||||u^{n+1} - v^{n+1}||$$

and by canceling out terms we get the contractive inequality

$$||u^{n+1} - v^{n+1}|| \leq ||u^n - v^n||$$

as desired. ∎

Proof of Theorem 4.3:

Proof. We assume that S is stable in the following sense:

$$||\hat{u}^{n+1}|| \leq ||\hat{u}^n||$$

for some l^p norm; common choices would be the l^2 or l^∞ norms. Because the shrink operator decreases the magnitude of each component of a vector, it will (strictly, because $\mu > 0$) decrease whatever norm is chosen (in fact, the shrink operator is a contraction in all l^p norms). It follows easily that

$$||\hat{u}_\mu^{n+1}|| \leq ||Q(\hat{u}_\mu^n, \ldots, \hat{u}_\mu^{n-k})|| \leq ||\hat{u}_\mu^n||$$

so that the stability of S implies the stability of S_μ. In fact S_μ is more stable than S.

The key observation for showing consistency of S_μ is that while $\text{shrink}(\cdot, \mu)$ is nonlinear, the amount of this nonlinearity is bounded. In particular,

$$\text{shrink}(x, \mu) = x + O(\mu)$$

for any x, with $|O(\mu)| \leq \mu$. Applying this observation to the definition of the sparse scheme and assuming (for the purpose of local truncation analysis) that both schemes have the same starting points $\hat{u}_\mu^n = \hat{u}^n, \ldots, \hat{u}_\mu^{n-k} = \hat{u}^{n-k}$,

$$\begin{aligned}
\hat{u}_\mu^{n+1} &= \text{shrink}(Q(\hat{u}_\mu^n, \ldots, \hat{u}_\mu^{n-k}), \mu) \\
&= Q(\hat{u}_\mu^n, \ldots, \hat{u}_\mu^{n-k}) + O(\mu) \\
&= Q(\hat{u}^n, \ldots, \hat{u}^{n-k}) + O(\mu) \\
&= \hat{u}^{n+1} + O(\mu).
\end{aligned}$$

This shows that locally, S and S_μ differ only by a $O(\mu)$ quantity. This quantity may naively be accounted as part of the local truncation error for the sparse scheme, in which case

$$\tau_\mu^n = \tau^n + O(\mu)$$

where τ^n denotes the local truncation error of S and τ_μ^n the local truncation error of S_μ.

For the consistency of S_μ, we need the local truncation error to be greater than first order; assuming the consistency of S and that $\mu = O(dt^{1+\delta})$ yields this result. When $\mu = O(dt^p)$ for some p such that $\tau^n = O(dt^p)$ as well, $\tau_\mu^n = O(dt^p)$ and the order of convergence of the scheme is unchanged. ∎

Proof of Theorem 4.4:

Proof. First, recall that the optimality condition for (3.2) is

$$\mu p(\hat{u}_\mu^{n+1}) + (I + dt\hat{L}_h)\hat{u}_\mu^{n+1} - \hat{u}_\mu^n + dt\hat{f}_h = 0 \tag{9.1}$$

where $p(\hat{u}_\mu^{n+1}) \in \partial\|\hat{u}_\mu^{n+1}\|_1$. For simplicity of notation, let $w := (\hat{u}_\mu^{n+1} - \hat{u}^{n+1})$. Assuming (again for the purpose of local truncation analysis) that both schemes have the same starting point $\hat{u}_\mu^n = \hat{u}^n$, subtracting the ordinary backward Euler update from this gives

$$(I + dt\hat{L}_h)w = \mu p(\hat{u}_\mu^{n+1})$$

which implies

$$\|(I + dt\hat{L}_h)w\|_\infty \leq \mu.$$

Then, using the fact that $\hat{L}_h$ is positive definite, we get

$$\frac{\|(I + dt\hat{L}_h)w\|_\infty}{\|w\|_2/N^{1/2}} \geq \frac{\|(I + dt\hat{L}_h)w\|_2}{\|w\|_2} \geq 1$$

which gives

$$\|w\|_{L^2(\Omega)} \sim \frac{\|w\|_2}{N^{1/2}} \leq \mu.$$

So, as with the explicit scheme,

$$\begin{aligned} & \hat{u}_\mu^{n+1} = \hat{u}^{n+1} + O(\mu) \qquad (\text{in } L^2(\Omega)) \\ \implies \quad & \tau_\mu^n = \tau^n + O(\mu) \end{aligned}$$

which yields consistency if $\mu = O(dt^{1+\delta})$ with $\delta > 0$, and implies the order of convergence is the same as that of the ordinary spectral scheme if $\mu = O(dt^p)$ with p such that $\tau^n = O(dt^p)$.

To prove stability of the scheme, return to (9.1) with $f = 0$ and take the inner product with $\hat{u}_\mu^{n+1}$ to get

$$\mu \|\hat{u}_\mu^{n+1}\|_1 + (\hat{u}_\mu^{n+1})^T (I + dt\hat{L}_h)\hat{u}_\mu^{n+1} - (\hat{u}_\mu^{n+1})^T \hat{u}_\mu^n = 0$$

which leads to

$$\begin{aligned} \|\hat{u}_\mu^{n+1}\|_2^2 &\le \langle \hat{u}_\mu^{n+1}, \hat{u}_\mu^n \rangle - \mu \|\hat{u}_\mu^{n+1}\|_1 - dt(\hat{u}_\mu^{n+1})^T \hat{L}_h \hat{u}_\mu^{n+1} \\ &\le \langle \hat{u}_\mu^{n+1}, \hat{u}_\mu^n \rangle \\ &\le \|\hat{u}_\mu^{n+1}\|_2 \|\hat{u}_\mu^n\|_2 \end{aligned}$$

and

$$\|\hat{u}_\mu^{n+1}\|_2 \le \|\hat{u}_\mu^n\|_2,$$

as desired. ∎

Proof of theorem 5.2:

Proof. We have

$$\begin{aligned} \frac{d}{dt}\frac{1}{2}||u(t,\cdot) - u_\delta(t,\cdot)||_2^2 &= \langle u - u_\delta, \partial_t u - \partial_t u_\delta \rangle \\ &= \langle u - u_\delta, -Lu + f - (-Lu_\delta + f - \delta\partial||\hat{u}(t)||_1) \rangle \\ &= -\langle u - u_\delta, Lu - Lu_\delta \rangle + \delta \langle u - u_\delta, \partial_0 ||\hat{u}_\delta||_1 \rangle \\ &\le \delta \langle u - u_\delta, \partial_0 ||\hat{u}_\delta||_1 \rangle \\ &\le \delta ||u - u_\epsilon||_2. \end{aligned}$$

It follows that

$$\frac{d}{dt}||u(t,\cdot) - u_\epsilon(t,\cdot)||_2 \le 2\delta$$

from which the result follows. ∎

Proof of theorem 7.1:

Proof. Let $\mathcal{F}_N[a(x/\epsilon)](k)$ denote the DFT of $a(x/\epsilon)$ on the grid; that is,

$$\mathcal{F}_N[a(x/\epsilon)](k) = \sum_{j=0}^{N-1} a\left(\frac{2\pi j}{N\epsilon}\right) e^{-2\pi i j k/N}.$$

Then

$$\begin{aligned}
\mathcal{F}_{2N}\left[a\left(\frac{x}{\epsilon/2}\right)\right](2k) &= \sum_{j=0}^{2N-1} a\left(\frac{2\pi j}{2N\cdot\epsilon/2}\right) e^{-2\pi i \frac{2k}{2N} j} \\
&= \sum_{j=0}^{N-1} a\left(\frac{2\pi j}{N\epsilon}\right)\left[e^{-2\pi i jk/N} + e^{-2\pi i (j+N)k/N}\right] \\
&= \sum_{j=0}^{N-1} a\left(\frac{2\pi j}{N\epsilon}\right)\left[e^{-2\pi i jk/N} + e^{-2\pi i jk/N} e^{-2\pi i k}\right] \\
&= 2\mathcal{F}_N[a(x/\epsilon)](k),
\end{aligned}$$

so that the even coefficients of $\mathcal{F}_{2N}\left[a\left(\frac{x}{\epsilon/2}\right)\right]$ are just those of $\mathcal{F}_N[a(x/\epsilon)]$. Also,

$$\begin{aligned}
\mathcal{F}_{2N}\left[a\left(\frac{x}{\epsilon/2}\right)\right](2k+1) &= \sum_{j=0}^{2N-1} a\left(\frac{2\pi j}{2N\cdot\epsilon/2}\right) e^{-2\pi i \frac{2k+1}{2N} j} \\
&= \sum_{j=0}^{N-1} a\left(\frac{2\pi j}{N\epsilon}\right)\left[e^{-2\pi i \frac{2k+1}{2N} j} + e^{-2\pi i \frac{2k+1}{2N}(j+N)}\right] \\
&= \sum_{j=0}^{N-1} a\left(\frac{2\pi j}{N\epsilon}\right) e^{-2\pi i \frac{2k+1}{2N} j}\left[1 + e^{-2\pi i \frac{2k+1}{2}}\right] \\
&= \sum_{j=0}^{N-1} a\left(\frac{2\pi j}{N\epsilon}\right) e^{-2\pi i \frac{2k+1}{2N} j}\left[1 + e^{-\pi i}\right] \\
&= 0
\end{aligned}$$

so that all odd coefficients vanish. These equalities give 7.3. ∎

REFERENCES

[1] Haim Brezis. *Monotone Operators, Nonlinear Semi-groups, and Applications*. Université Pierre et Marie Curie, Laboratoire d'Analyse Numérique, 1974.

[2] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[3] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1998.

[4] Michael G Crandall and Thomas M Liggett. Generation of semi-groups of nonlinear transformations on general banach spaces. *American Journal of Mathematics*, 93(2):265–298, 1971.

[5] Ingrid Daubechies, Olof Runborg, and Jing Zou. A sparse spectral method for homogenization multiscale problems. *Multiscale Modeling & Simulation*, 6(3):711–740, 2007.

[6] David L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995.

[7] Lawrence C Evans. *Partial differential equations*. 1998.

[8] Thomas Y Hou and Xiao-Hui Wu. A multiscale finite element method for elliptic problems in composite materials and porous media. *Journal of computational physics*, 134(1):169–189, 1997.

[9] Ioannis G Kevrekidis, C William Gear, James M Hyman, Panagiotis G Kevrekidid, Olof Runborg, Constantinos Theodoropoulos, et al. Equation-free, coarse-grained multiscale computation: Enabling mocroscopic simulators to perform system-level analysis. *Communications in Mathematical Sciences*, 1(4):715–762, 2003.

[10] James Nolen, George Papanicolaou, and Olivier Pironneau. A framework for adaptive multiscale methods for elliptic problems. *Multiscale Modeling & Simulation*, 7(1):171–196, 2008.
[11] G Papanicolau, A Bensoussan, and J-L Lions. *Asymptotic analysis for periodic structures*. North Holland, 1978.
[12] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.
[13] Hayden Schaeffer, Russel Caflisch, Cory Hauck, and Stanley Osher. Sparse dynamics for partial differential equations. *Proceedings of the National Academy of Sciences*, 110(17):6634–6639, 2013.
[14] Weinan E, Bjorn Engquist, and Zhongyi Huang. Heterogeneous multiscale method: a general methodology for multiscale modeling. *Physical Review B*, 67(9):092101, 2003.
[15] Weinan E, Bjorn Engquist, Xiantao Li, Weiqing Ren, and Eric Vanden-Eijnden. Heterogeneous multiscale methods: a review. *Communications in computational physics* 2, no. 3 (2007): 367-450.
[16] Emmanuel J Candes, Yonina C Eldar, Thomas Strohmer, and Vladislav Voroninski. Phase retrieval via matrix completion. *SIAM Journal on Imaging Sciences*, 6(1):199–225, 2013.
[17] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.
[18] David L Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.
[19] Tom Goldstein and Stanley Osher. The split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.
[20] Jian-Feng Cai, Stanley Osher, and Zuowei Shen. Linearized bregman iterations for compressed sensing. *Mathematics of Computation*, 78(267):1515–1536, 2009.
[21] Russel E Caflisch, Stanley J Osher, Hayden Schaeffer, and Giang Tran. PDEs with compressed solutions. *arXiv preprint arXiv:1311.5850*, 2013.
[22] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.
[23] Lance J Nelson, Gus LW Hart, Fei Zhou, and Vidvuds Ozoliņš. Compressive sensing as a paradigm for building physics models. *Physical Review B*, 87(3):035125, 2013.
[24] Vidvuds Ozoliņš, Rongjie Lai, Russel Caflisch, and Stanley Osher. Compressed modes for variational problems in mathematics and physics. *Proceedings of the National Academy of Sciences*, 110(46):18368–18373, 2013.
[25] Vidvuds Ozolinš, Rongjie Lai, Russel Caflisch, and Stanley Osher. Compressed plane waves-compactly supported multiresolution basis for the laplace operator. *Proceedings of the National Academy of Sciences*, 2013.
[26] Thomas Y Hou, Zuoqiang Shi, and Peyman Tavallali. Sparse time frequency representations and dynamical systems. *arXiv preprint arXiv:1312.0202*, 2013.
[27] Steven L Brunton, Jonathan H Tu, Ido Bright, and J Nathan Kutz. Compressive sensing and low-rank libraries for classification of bifurcation regimes in nonlinear dynamical systems. *arXiv preprint arXiv:1312.4221*, 2013.
[28] Haim Brezis. Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert *Elsevier*, 5, 1973.
[29] Yalchin Efendiev and Thomas Y Hou. *Multiscale finite element methods: theory and applications*. Vol. 4. Springer, 2009.
[30] Haim Brezis. Solutions with compact support of variational inequalities. *Russian Mathematical Surveys*, 29(2):103–108, 1974.