Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 224 (2007) 1243–1254
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Locally-corrected spectral methods and overdetermined elliptic systems

John Strain *

Department of Mathematics, University of California, 970 Evans Hall #3840, Berkeley, CA 94720-3840, United States

Received 17 April 2006; received in revised form 12 November 2006; accepted 13 November 2006
Available online 8 January 2007

Abstract

We present fast locally-corrected spectral methods for linear constant-coefficient elliptic systems of partial differential equations in d-dimensional periodic geometry. First, arbitrary second-order elliptic systems are converted to overdetermined first-order systems. Overdetermination preserves ellipticity, while first-order systems eliminate mixed derivatives, resolve convection–diffusion conflicts, and simplify derivative computations. Second, a periodic fundamental solution is derived by Fourier analysis and mollified for rapid convergence, independent of the regularity of the elliptic problem. Third, a new Ewald summation technique for first-order elliptic systems locally corrects the mollified solution to achieve high-order accuracy. We also discuss second-kind boundary integral equations based on single layer potentials formed with the mollified and corrected fundamental solution, which form a useful toolkit for solving general elliptic boundary value problems in general domains. The resulting spectral methods provide highly accurate solutions and derivatives for periodic problems.

PACS: 02.30.Jr; 02.60.−x; 02.30.Nw; 46.15.−x; 47.11.−j; 02.70.−c

Keywords: Spectral methods; Elliptic systems; Fourier series; Ewald summation; Local correction; Moving interface problems; Boundary integral method

1. Introduction

A wide variety of time-independent physical problems find mathematical expression as second-order linear constant-coefficient elliptic systems

$$\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{l=1}^{s} a_{ij}^{kl} u_{,ij}^{l} + \sum_{j=1}^{d}\sum_{l=1}^{s} b_{j}^{kl} u_{,j}^{l} + \sum_{l=1}^{s} c^{kl} u^{l} = f^{k}, \quad 1 \leqslant k \leqslant s, \tag{1}$$

* Tel./fax: +1 510 642 8204.
E-mail address: strain@math.berkeley.edu.
URL: http://math.berkeley.edu/~strain.

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2006.11.017

where $u^l_{,j}$ is the partial derivative of u^l with respect to x_j. Such systems include the Poisson, Stokes and linear elasticity equations, which are often solved by specialized, inflexible codes for specific systems [1–3]. In this paper, we present a flexible new top-down approach which solves a wide spectrum of elliptic systems with uniform efficiency, and apply our new approach to develop accurate and efficient new spectral methods for elliptic problems in periodic domains. The new methods promise uniform accuracy for nonsmooth solutions and complex domains which are inaccessible to classical Fourier techniques.

Our approach converts any system (1) to a simple overdetermined first-order system

$$\sum_{j=1}^{d} A_j u_{,j} + A_0 u = f,$$

where each A_j is a $p \times q$ matrix and u is a q-vector. The conversion eliminates mixed derivatives, resolves convection-diffusion conflicts, and reduces condition numbers from $O(N^2)$ to $O(N)$ at resolution N. It solves all elliptic systems with a single efficient code, because linear algebra takes its proper place: correlating local relations between solution components. In the context of boundary integral formulations, the first-order conversion eliminates complicated relations between higher-order potential operators and employs single-layer potentials exclusively.

The paper is organized as follows. In Section 2, we convert arbitrary second-order elliptic problems to overdetermined first-order systems. In Section 3, we represent the solution to an overdetermined periodic first-order system as a "box potential" computed by integration against a fundamental solution. A periodic fundamental solution is derived by Fourier analysis in Section 4. A natural definition of ellipticity for first-order systems is justified. Suboptimal convergence of standard spectral methods for problems with nonsmooth solutions is discussed in Section 5. The classical Ewald summation technique which resolves convergence difficulties for the Poisson equation is reviewed in Section 6. In Section 7, a new Ewald summation technique for first-order elliptic systems is presented. It splits the fundamental solution into a global rapidly-converging Fourier series, mollified by a matrix exponential, and an error term. In classical Ewald summation, the error term is computed via special functions and integration, which cannot easily be done for a general elliptic system. Instead, we compute the error term by a simple Taylor expansion in Fourier space, which locally corrects the mollified fundamental solution by an asymptotic series of local differential operators. Our new mollification and local correction techniques are combined with the fast Fourier transform, Padé codes for small dense matrix exponentials, and high-order uncentered differencing to solve first-order elliptic systems in Section 8. In Section 9, we present a simple algebraic algorithm for the automatic computation of local correction coefficients which achieve high-order accuracy at minimal cost. Section 10 presents numerical experiments which verify efficiency and accuracy. In Section 11 we discuss extensions such as boundary integral equations for complex domains and variable-coefficient systems.

2. Conversion to first-order systems

Conversion to a first-order system replaces tiresome case-by-case analyses by linear algebra, computes derivatives of the solution automatically, and fosters the development of practical yet general codes for elliptic systems. In previous work on moving interfaces [4–6], for example, the various physical models of bulk processes require a wide array of solvers for elliptic and parabolic problems, and move the interface via computed normal derivatives of the solution. This complicated and sensitive technology would be greatly simplified by efficient codes for the stable computation of solutions and derivatives to general elliptic systems.

We convert the second-order system (1) to a first-order system by introducing all solution components u^l and their first derivatives $u^l_{,j}$ as components of a q-vector $u = (u^1, u^2, \ldots, u^s, u^1_{,1}, u^2_{,1}, \ldots, u^s_{,d}) = (u_1, u_2, \ldots, u_q) \in \mathbb{R}^q$. The vector u satisfies $p = (d + d(d-1)/2 + 1)s \geqslant q = (1 + d)s$ equations, which guarantee the following three conditions:

(a) the first s components $(u_1, u_2, \ldots, u_s)$ constitute a solution to the original second-order elliptic system in the new variables

$$\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{l=1}^{s} a_{ijkl}u_{l+si,j} + \sum_{j=1}^{d}\sum_{l=1}^{s} b_{jkl}u_{l+sj} + \sum_{l=1}^{s} c_{kl}u_l = f^k, \quad 1 \leqslant k \leqslant s,$$

(b) the subsequent ds components are the appropriate derivatives of the first s

$$u_{i,j} = u_{i+sj}, \quad 1 \leqslant i \leqslant s, \quad 1 \leqslant j \leqslant d,$$

(c) the $d(d-1)s$ mixed partial derivatives are equal, so the appropriate u_j are the derivatives of a single function u^l

$$u_{i+sk,j} = u_{i+sj,k}, \quad 1 \leqslant i \leqslant s, \quad 1 \leqslant k < j \leqslant d.$$

Conditions (a)–(c) can be summarized by a first-order linear system with matrix coefficients A_j

$$Au := \sum_{j=1}^{d} A_j u_{,j} + A_0 u = f. \tag{2}$$

The matrices A_j have p rows and $q \leqslant p$ columns, so the system appears overdetermined. Nonetheless, if an appropriate definition of ellipticity is satisfied (Section 4), the boundary value problem is Fredholm: solutions exist for data f subject to a finite number of compatibility conditions, and are unique up to a finite-dimensional kernel [7–9]. A complete algebraic theory of such conversions, and of the ellipticity of the resulting systems, is developed in [10].

A similar conversion is employed in the first-order system least squares (FOSLS) approach, which solves the resulting first-order system by a finite element method [11]. The FOSLS approach treats conditions (a), (b) and (c) differently, as in the Agmon–Douglis–Nirenberg theory of elliptic systems [12]. Formulation of second-order equations as first-order symmetric hyperbolic systems has also been standard in the analysis and numerical solution of hyperbolic equations for wave propagation [13–15].

Conversion to first-order systems is straightforward for the following standard physical problems

Example 1. For the Poisson equation $\Delta u = f$ in dimension $d = 2$, an equivalent first-order system is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} u \\ u_{,x} \\ u_{,y} \end{bmatrix}_{,x} + \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ u_{,x} \\ u_{,y} \end{bmatrix}_{,y} + \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ u_{,x} \\ u_{,y} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ f \end{bmatrix}.$$

Example 2. For the three-dimensional steady incompressible Stokes equations

$$-v\Delta u + \nabla P = f, \quad \nabla \cdot u = 0,$$

the pressure P satisfies a first-order equation, so conversion yields $3 + 1 + 9 + 9 = 22$ first-order equations in $3 + 1 + 3 \times 3 = 13$ unknowns $(u_1, u_2, u_3, p, u_{1,1}, \ldots, u_{3,3})$.

Example 3. The time-harmonic Maxwell system

$$\nabla \times E - \mathrm{i}\omega\mu H = 0, \quad \nabla \cdot E = \rho, \quad \nabla \times H + \mathrm{i}\omega\epsilon E = J, \quad \nabla \cdot H = 0, \tag{3}$$

is naturally posed in first-order overdetermined form, so conversion is unnecessary.

3. Integral formulas for solutions of first-order systems

We represent any smooth solution u, of a first-order system $Au = f$, as a "box potential" $u = Bf$ formed by integration of the right-hand-side f against the fundamental solution φ constructed in Section 4. This representation is analogous to the solution of a linear system $Ax = b$ by the inverse matrix $A^{-1}b$.

By Gauss' divergence theorem [13], any smooth function g satisfies

$$\int_Q g_{,j}\,\mathrm{d}x = \int_\Gamma n_j g\,\mathrm{d}\gamma, \quad 1 \leqslant j \leqslant d,$$

where $n = (n_1, n_2, \ldots, n_d)$ is the outward unit normal vector to the boundary Γ of a smooth d-dimensional domain Q. Suppose we can find a $q \times p$ matrix-valued "fundamental solution" $\varphi = \varphi_x : \mathbb{R}^d \to \mathbb{R}^{q\times p}$ with pole at $x \in Q$, a distributional solution of the inhomogeneous adjoint system

$$\sum_{j=1}^{d} -\varphi_{,j}A_j + \varphi A_0 = \delta_x I. \tag{4}$$

Here δ_x is a Dirac delta at x and I is the $q \times q$ identity matrix. Apply Gauss' divergence theorem to each product $\varphi A_j u$ and use the product rule for differentiation

$$\int_Q \varphi_{,j}A_j u\,\mathrm{d}x + \int_Q \varphi A_j u_{,j}\,\mathrm{d}x = \int_\Gamma n_j \varphi A_j u\,\mathrm{d}\sigma.$$

Sum over $j = 1$ to d, add and subtract the zero-order term in the first-order system to get

$$\int_Q \sum_{j=1}^{d} \varphi_{,j}A_j u - \varphi A_0 u\,\mathrm{d}x + \int_Q \sum_{j=1}^{d} \varphi A_j u_{,j} + \varphi A_0 u\,\mathrm{d}x = \int_\Gamma \sum_{j=1}^{d} n_j \varphi A_j u\,\mathrm{d}\sigma. \tag{5}$$

In general, Eq. (5) leads to a boundary integral equation for u. For the present periodic problem, we specialize to a d-dimensional cube $Q = [a,b]^d$ and impose periodic boundary conditions on u and φ_x. Then the boundary term vanishes by periodicity and $u(x)$ is determined everywhere in Q as the "box potential" Bf of the right-hand-side f

$$u(x) = Bf(x) := \int_Q \varphi_x(y) f(y)\,\mathrm{d}y.$$

4. The Fourier series of a fundamental solution

Next we construct an effective evaluation formula for the fundamental solution φ. Assume by scaling if necessary that the cube $Q = [-\pi, \pi]^d \subset \mathbb{R}^d$. Then the standard Fourier series pair on Q reads

$$f(x) = \sum_{k\in\mathbb{Z}^d} \hat{f}(k)\mathrm{e}^{-\mathrm{i}k^{\mathrm{T}}x}, \quad \hat{f}(k) = \frac{1}{|Q|}\int_Q f(y)\mathrm{e}^{\mathrm{i}k^{\mathrm{T}}y}\,\mathrm{d}y, \quad \widehat{f_{,j}}(k) = -\mathrm{i}k_j\hat{f}(k),$$

where $k = (k_1, k_2, \ldots, k_d) \in \mathbb{Z}^d$ means each entry k_j of the d-vector k is a positive or negative integer or zero, $|Q| = (2\pi)^d$ is the volume of Q and $\mathrm{d}y = \mathrm{d}y_1\mathrm{d}y_2 \ldots \mathrm{d}y_d$. Thus by Eq. (4), the Fourier coefficients of φ_x are $q \times p$ matrices satisfying

$$\widehat{\varphi_x}(k)\left(\sum_{j=1}^{d} \mathrm{i}k_j A_j + A_0\right) = \widehat{\varphi_x}(k)A(k) = \frac{\mathrm{e}^{\mathrm{i}k^{\mathrm{T}}x}}{|Q|} I, \tag{6}$$

where $A = A(k) = \mathrm{i}\sum_{j=1}^{d} k_j A_j + A_0$ is $p \times q$ with $q \leqslant p$ and I is the $q \times q$ identity matrix. A solution $\widehat{\varphi_x}$ exists if A is injective, meaning it has linearly independent columns, maximal rank q, or equivalently a trivial nullspace. This observation justifies a classical definition of ellipticity (in the sense of Protter [7–10]), which ensures that $\widehat{\varphi_x}$ exists for almost all k:

Definition 1. The first-order system Eq. (2) is elliptic if the "principal part" $\mathrm{i}\sum_{j=1}^{d} k_j A_j$ is injective for every nonzero vector k.

This definition implies that any elliptic system on a periodic box is solvable in Fourier space for almost every k, because the zero-order coefficient A_0 is always small relative to the principal part for all sufficiently large k and the set of injective $p \times q$ matrices is open. In real space, the exceptional k's correspond to a finite-dimensional space of constraints on the right-hand-side and a finite-dimensional nullspace. An elliptic system is therefore "Fredholm" [16] or "normally solvable".

A solution of the underdetermined linear system (6) for $\widehat{\varphi_x}$ can be explicitly found for almost every k by the pseudoinverse formula

$$\widehat{\varphi_x}(k) = \frac{e^{ik^T x}}{|Q|}(A^*(k)A(k))^{-1}A^*(k) = \frac{e^{ik^T x}}{|Q|}A^\dagger(k),$$

where $A^\dagger$ is the Moore-Penrose generalized inverse [17]. Thus the fundamental solution has a Fourier series

$$\varphi_x(y) = \frac{1}{|Q|}\sum_{k\in\mathbb{Z}^d} e^{ik^T(x-y)}(S(k))^{-1}A^*(k), \tag{7}$$

where $S = A^*A$, if $A(k)$ is injective for all $k \in \mathbb{Z}^d$. Isolated k vectors where injectivity fails correspond to compatibility conditions such as mean-zero requirements, and must be treated carefully in practical computations. We verify ellipticity for two of the systems from Section 2:

Example 4. For the 2D Poisson equation, everything can be worked out explicitly

$$A = \begin{bmatrix} ik_1 & -1 & 0 \\ ik_2 & 0 & -1 \\ 0 & -ik_2 & ik_1 \\ 0 & ik_1 & ik_2 \end{bmatrix}, \quad A^*A = \begin{bmatrix} k^2 & ik_1 & ik_2 \\ -ik_1 & 1+k^2 & 0 \\ -ik_2 & 0 & 1+k^2 \end{bmatrix},$$

where $k^2 = k_1^2 + k_2^2 = |k|^2$. and the inverse matrix is

$$(A^*A)^{-1} = \frac{1}{k^4(1+k^2)}\begin{bmatrix} (1+k^2)^2 & ik_1(1+k^2) & ik_2(1+k^2) \\ -ik_1(1+k^2) & k^2(1+k^2)-k_2^2 & k_1k_2 \\ -ik_2(1+k^2) & k_1k_2 & k^2(1+k^2)-k_1^2 \end{bmatrix},$$

Since the principal part

$$i\sum_{j=1}^{d} k_jA_j = \begin{bmatrix} ik_1 & 0 & 0 \\ ik_2 & 0 & 0 \\ 0 & -ik_2 & ik_1 \\ 0 & ik_1 & ik_2 \end{bmatrix},$$

is injective for $k \neq 0$, the first-order system is elliptic. It would not be elliptic if we omitted the third row, which requires the equality of the mixed partial derivatives. Thus overdetermination preserves ellipticity.

Example 5. The time-harmonic Maxwell system (3) is elliptic by Definition 4. Indeed, if $k \times E = 0$ and $k \cdot E = 0$ then components of E which are perpendicular and parallel to k both vanish, so E must vanish. The overdetermined $p = 8 \times 6 = q$ system (3) avoids well-known difficulties in treating the parallel conditions on $\nabla \cdot H$ and $\nabla \cdot E$ as auxiliary constraints [18].

5. Divergence issues

Given the Fourier series representation (7) of the fundamental solution, we could solve an elliptic system (2), with a smooth solution u in the cube Q, by the following Fourier method: Approximate N^d Fourier coefficients $\hat{f}(k)$ by the trapezoidal rule, evaluated with the Fast Fourier Transform (FFT), multiply each coefficient $\hat{f}(k)$ by the corresponding $\widehat{\varphi_0}(k)$, and evaluate the Fourier series for u on a regular grid with another FFT. However, the Fourier series representation (7) of a fundamental solution usually diverges in any C^r norm, because $\widehat{\varphi}(k) = O(1/|k|)$ as $k \to \infty$. The divergence is more severe in higher dimensions where $\sum O(1/|k|) = O(|k|^{d-1})$. Thus the Fourier method depends on smooth solutions for rapid convergence, and diverges in C^r norms via the Gibbs phenomenon when solutions are not smooth [19,20].

We use Ewald summation to resolve divergence issues and improve the accuracy of the computed solution. It multiplies each Fourier coefficient by a carefully structured Gaussian filter to achieve rapid convergence,

and then compensates by a local correction of the filtering error. We begin with a review of the classical paradigm in the next section, and derive a new Ewald summation technique for elliptic systems in Sections 7–9.

6. Classical Ewald summation

The classical version of Ewald summation [21] is widely used in computational chemistry [22,23] and fluid mechanics [24,25], and has been used to construct fast Poisson and Stokes solvers [26–28]. It separates global from local effects to provide a rapidly converging formula for the periodic mean-zero fundamental solution $\mathscr{G}$ of the Poisson equation.

A simple derivation of classical Ewald summation can be based on Fourier analysis and the method of images, for the periodic fundamental solution $\mathscr{K}_t$ of the heat equation $u_{,t} = \Delta u$, given by two separate but equal formulas

$$\mathscr{K}_t(x,t) = \frac{1}{|Q|}\sum_k e^{-t|k|^2} e^{ik^T x} = (4\pi t)^{-d/2}\sum_k e^{-|x-2\pi k|^2/4t}. \tag{8}$$

Correspondingly, the fundamental solution $\mathscr{G}$ of the Poisson equation has a Fourier series

$$\mathscr{G}(x) = \frac{1}{|Q|}\sum_{k\neq 0}\frac{-1}{|k|^2} e^{ik^T x} = \frac{-1}{|Q|}\sum_{k\neq 0}\int_0^\infty e^{-t|k|^2}\,dt\; e^{ik^T x}.$$

Splitting the time integral at $t = \tau$, recognizing the heat kernel $\mathscr{K}_t$, and using each formula from Eq. (8) where it converges the fastest, expresses $\mathscr{G}$ as the sum of two rapidly-converging series

$$\begin{aligned}\mathscr{G}(x) &= \frac{-1}{|Q|}\sum_{k\neq 0}\frac{e^{-\tau|k|^2}}{|k|^2} e^{ik^T x} + \int_0^\tau \left(\frac{1}{|Q|} - (4\pi t)^{-d/2}\sum_k e^{-|x-2\pi k|^2/4t}\,dt\right)\\ &= \frac{1}{|Q|}\left(\tau - \sum_{k\neq 0}\frac{e^{-\tau|k|^2}}{|k|^2} e^{ik^T x}\right) - \sum_k \int_0^\tau (4\pi t)^{-d/2} e^{-|x-2\pi k|^2/4t}\,dt.\end{aligned} \tag{9}$$

The remaining time integrals can be evaluated exactly in terms of incomplete gamma functions, which account for the singularity at $x = 0$ and decay rapidly as $|x|$ increases. By optimizing the choice of τ, the fundamental solution can be evaluated with less than a dozen terms of each series. Combined with nonuniform fast Fourier transforms (FFTs), this yields fast solvers for various problems associated with the Poisson equation [26,27]. The Stokes equation is equivalent to the Poisson and biharmonic operators, and complicated by the divergence constraint [24,25,28]. However, Ewald summation has not been derived for systems such as linear elasticity, because exact evaluation of the integral from 0 to τ in Eq. (9) requires a matrix-valued special function technology which does not exist. Our new Ewald summation schemes are designed for general first-order elliptic systems.

7. Ewald summation for elliptic systems

Ewald summation for a general first-order elliptic system relies on two key observations.

First, the Hermitian matrix $S(k) = A^*(k)A(k)$ is positive definite for an elliptic system. This follows from injectivity of $A(k)$ and implies that

$$\begin{aligned}S(k)^{-1} &= \int_0^\infty e^{-tS(k)}\,dt\\ &= \int_0^\tau e^{-tS(k)}\,dt + \int_\tau^\infty e^{-tS(k)}\,dt\\ &= (I - e^{-\tau S(k)})S(k)^{-1} + e^{-\tau S(k)}S(k)^{-1},\end{aligned} \tag{10}$$

where e^{tS} denotes the matrix exponential, defined for Hermitian matrices S by unitary diagonalization

$$e^{tS} = e^{tU\Lambda U^*} = Ue^{t\Lambda}U^* \quad \text{where } S = U\Lambda U^*, \quad UU^* = U^*U = I.$$

Thus the natural mollifier $e^{-\tau S}$ splits the time integral (10) at $t = \tau$, as in the classical Ewald summation (9), and expresses the fundamental solution as a rapidly converging global contribution φ^{F} and a local correction φ^{L}

$$\begin{aligned}\varphi_x(y) =& \frac{1}{|Q|}\sum_k e^{-\tau S(k)} S(k)^{-1} A^*(k) e^{ik^{\mathrm{T}}(x-y)} + \frac{1}{|Q|}\sum_k (I - e^{-\tau S(k)}) S(k)^{-1} A^*(k) e^{ik^{\mathrm{T}}(x-y)} \\ =& \varphi_x^{\mathrm{F}}(y) + \varphi_x^{\mathrm{L}}(y).\end{aligned}$$

Second, the local contribution φ^{L}, which does not converge rapidly as a Fourier series, can conveniently be evaluated by an asymptotic series of real-space differential operators. This contrasts with classical Ewald summation, where the real-space kernel of the local contribution is expressed by special functions. General elliptic systems do not permit such expression, because the appropriate matrix-valued special functions are not available. Instead, we employ a Taylor series expansion in Fourier space. Because the mollifier is a matrix exponential $e^{-\tau S}$, this eliminates the nonlocal operator S^{-1} and yields a series of local differential operators

$$(I - e^{-\tau S}) S^{-1} A^* = \left(\tau - \frac{1}{2}\tau^2 S + \frac{1}{6}\tau^3 S^2 - \cdots\right) A^*. \tag{11}$$

For a small number of k values, $A(k)$ may not be injective. For those k, the pseudoinverse matrix $S^{\dagger}$ replaces $S^{-1}A^*$ and $SS^{\dagger}$ becomes the projection P_0 perpendicular to the kernel of $A(k)$, rather than the identity. The resulting modified Eq. (11) corresponds to the τ term in Eq. (9) for the classical Poisson case.

After Taylor expansion, the local part of the fundamental solution has a real-space asymptotic expansion in terms of the usual Dirac point mass

$$\delta_x(y) = \frac{1}{|Q|}\sum_k e^{ik^{\mathrm{T}}(x-y)},$$

and its derivatives

$$\left(\tau - \frac{1}{2!}\tau^2 S + \frac{1}{3!}\tau^3 S^2 - \cdots\right) A^* \delta_x(y) I. \tag{12}$$

Here the differential operator A, its formal adjoint A^* and square S are given by

$$Au = A_j u_{,j} + A_0 u, \quad A^* v = -A_j^* v_{,j} + A_0^* v, \quad S = A^* A.$$

As the most immediate consequence, the simplest local differential correction $\tau A^* \delta_x$ improves the order of accuracy of the mollified fundamental solution from $O(\tau)$ to $O(\tau^2)$.

We now translate our Ewald summation technique from the fundamental solution to a specific solution u of the elliptic system.

8. Fourier solution of the elliptic system

Our Ewald summation technique for the fundamental solution φ splits the box potential $u = Bf$ into global and local parts as follows:

$$\begin{aligned}u(x) =& B_{\mathrm{F}} f(x) + B_{\mathrm{L}} f(x) \\ =& \sum_{k\in\mathbb{Z}^d} e^{-\tau S} (S(k))^{-1} A^*(k) \hat{f}(k) e^{-ik^{\mathrm{T}}x} + \left(\tau - \frac{1}{2!}\tau^2 S + \frac{1}{3!}\tau^3 S^2 - \cdots\right) A^* f(x).\end{aligned}$$

$B_{\mathrm{F}} f$ is a rapidly converging Fourier series, while $B_{\mathrm{L}} f$ is a local correction. Each can be evaluated by standard tools of numerical analysis, tailored to their specific properties.

We evaluate $B_{\mathrm{F}} f$ on a regular grid with mesh size $h = 2\pi/N$ in three steps. First, the Fourier coefficients $\hat{f}(k)$ of f are approximated by the trapezoidal rule and efficiently evaluated with the FFT. If f is smooth and

periodic, the error in $\hat{f}(k)$ is spectrally small. Otherwise, the trapezoidal rule will give less accurate results; then specialized techniques such as attenuation factors [29], or the piecewise-polynomial nonuniform FFT of [30], compute the Fourier coefficients of f with uniform error $O(h^p)$ in optimal time. Second, the matrix exponentials $e^{-\tau S(k)}$ which mollify the Fourier series are evaluated by standard Padé approximation codes such as `padm` from `expokit` [31]. They can be efficiently precomputed, stored and reused if the same elliptic system is to be solved repeatedly with different data, as in time stepping. $S(k)^{-1}$ is applied by standard Cholesky decomposition or the SVD [32]. Mean value conditions may be imposed by the SVD if necessary. Third, $B_F f$ is evaluated on the grid with another FFT.

There are two sources of error in the computation of $B_F f$, the spectral or $O(h^p)$ error in the Fourier coefficients of f, and the Fourier series truncation error $O(e^{-\tau N^2})\|f\|_1$. It is worth noting that both errors can be tightly controlled even if f is a discontinuous function, a measure or a distribution.

9. High-order local correction

We approximate the local correction $B_L f$ on the same uniform grid, using the same grid values of f. High-order accuracy is obtained with uncentered finite differencing based on polynomial interpolation.

Conceptually, we can replace $f(x)$ by a Lagrange interpolation polynomial P based on a square stencil containing $(2s+1)^d$ grid neighbors of each evaluation point x. Then an exact algebraic computation of the first m terms

$$B_L^m P(x) = \left(\tau - \frac{1}{2!}\tau^2 S + \frac{1}{3!}\tau^3 S^2 - \cdots \pm \frac{1}{m!}\tau^m S^{m-1}\right) A^* P(x), \tag{13}$$

of the spectral Taylor series for $B_L P(x)$ will yield accuracy of order $B_L f(x) - B_L^m P(x) = O(\tau^{m+1} + \tau h^{2s} + \tau^2 h^{2s-2} + \cdots)$. The $O(\tau^{m+1})$ term is due to truncating the asymptotic series, while the other terms are due to polynomial interpolation of f. The extra factors of τ comes from the $O(\tau)$ and higher size of the local correction terms.

The simplest example of this concept, with $m = s = 1$, makes $B_L^1 = \tau A^*$ exact for multilinear polynomials P. It truncates the local correction to

$$B_L^1 f(x) := \tau A^* f(x) = \tau\left(\sum_{j=1}^{d} -A_j^* f_{,j}(x) + A_0^* f(x)\right),$$

and replaces the derivatives $f_{,j}(x)$ with difference approximations built from polynomial interpolation at the nearest neighbors of x. The error is the sum of an $O(\tau^2)$ series truncation error, and the $O(\tau h^2)$ error due to replacing $f_{,j}(x)$ by the corresponding derivative $P_{,j}(x)$ of a multilinear interpolating polynomial P. The natural choice $\tau = O(h^2)$, which keeps the Fourier series mollification error below level $\epsilon = O(e^{-\tau N^2})$, gives fourth-order accuracy if f has bounded third-order derivatives.

However, the idea of interpolating and evaluating is only a conceptual tool. It shows that there are $q \times p$ matrix weights w_{ij} which make the following formula exact for polynomials P up to a certain degree

$$B_L^m P(x_i) = \sum_j w_{ij} P(x_j).$$

The most efficient approach to determining the weights w_{ij} would solve a linear system (underdetermined for stability [33]), which requires exactness for some stable basis of the space of polynomials. This is cumbersome to program, so we employ the following convenient approach.

We base a general evaluation scheme for $B_L^m P(x)$ on high-order equidistant uncentered finite difference stencils for the first derivatives

$$g'(x + ih) = \sum_{j=-s}^{s} c_{ij} g(x + jh) + O(h^{2s}).$$

Such stencils, generated by Fornberg's standard technique [34], yield tensor coefficients which produce Af or A^*f to order $O(h^{2s})$ everywhere on the stencil $\sigma_i = \{x_j \| x_i - x_j| \leqslant sh\}$ from values of f on σ_i. Matrix multiplication and addition then build an $O(\tau^{m+1} + \tau h^{2s})$-accurate stencil for the local correction.

The algorithm is detailed in the following pseudocode:

Algorithm 1. Compute high-order local correction coefficient matrix $\mathscr{B}$:

Compute $K = 2s + 1$ by K matrix C, by Fornberg's method [34], such that
$P'(ih) = \frac{1}{h}\sum_{j=-s}^{s} c_{ij} P(jh)$
is exact for $|i| \leqslant s$ for all polynomials P of degree $\leqslant K - 1$ in each variable.
Compute K^d by K^d matrices $\mathscr{A}$ of $p \times q$ blocks and $\mathscr{A}^*$ of $q \times p$ blocks by
$\mathscr{A}_{\alpha\beta} = (\sum_{j=1}^{d} A_j c_{\alpha_j\beta_j} \prod_{i\neq j} \delta_{\alpha_i\beta_i}) + A_0\delta_{\alpha\beta}$
$\mathscr{A}^*_{\alpha\beta} = (-\sum_{j=1}^{d} A^*_j c_{\alpha_j\beta_j} \prod_{i\neq j} \delta_{\alpha_i\beta_i}) + A^*_0\delta_{\alpha\beta}$
Here $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)$ and β are multiindices of integers in $[-s,s]$.
Initialize matrices $\mathscr{B}$ and term $\mathscr{T}$ of the same size as $\mathscr{A}^*$ by
$\mathscr{B} = \mathscr{T} = \mathscr{A}^*$
do $\mu = 2 \ldots m$
$\mathscr{T} \leftarrow -\tau \mathscr{A}\mathscr{A}^*\mathscr{T}/\mu$
$\mathscr{B} \leftarrow \mathscr{B} + \mathscr{T}$
end do

An even more accurate version of the algorithm replaces B^m_{L} by the exact operator stencil $\mathscr{E}(\mathscr{A}) = (\mathrm{e}^{-\tau\mathscr{A}\mathscr{A}^*} - I)(\mathscr{A}^*\mathscr{A})^{-1}\mathscr{A}^*$ to obtain error $\mathrm{O}(\tau h^{2s})$.

10. Numerical results

We implemented a 2D version of our algorithm in the C programming language and verified its accuracy and efficiency on a gallery of test cases. For a fixed random C^4 solution (Fig. 1)

$$u(x) = \sum_{k=1}^{\infty} r_k k^{-6} \cos(kx_1)\cos((k+1)x_2), \qquad \text{random} \quad r_k \in [-1,1]$$

of the Poisson equation $\Delta u = f$, solved as a 4×3 first-order system for $u_1 = u$, $u_2 = u_{,x}$ and $u_3 = u_{,y}$, Fig. 2 exhibits high-order convergence. Maximum-norm errors E, in the $3N^2$-vector consisting of u and its first derivatives evaluated on the grid, are plotted in a log–log plot vs. CPU seconds T as four parameters vary: grid size N, Fourier cutoff τN^2, number of terms m, and local correction order $2s$. The lower envelope of this cloud of results exhibits high-order convergence: $E = \mathrm{O}(T^{-4}) = \mathrm{O}(N^{-8}) = \mathrm{O}(h^8)$ as $N \to \infty$. Naturally this lower envelope consists of cases where the order of accuracy varies with the accuracy desired: Low accuracy is more efficiently obtained with small m, s and N, while a eighth-order scheme with 5 order-8 corrections produces

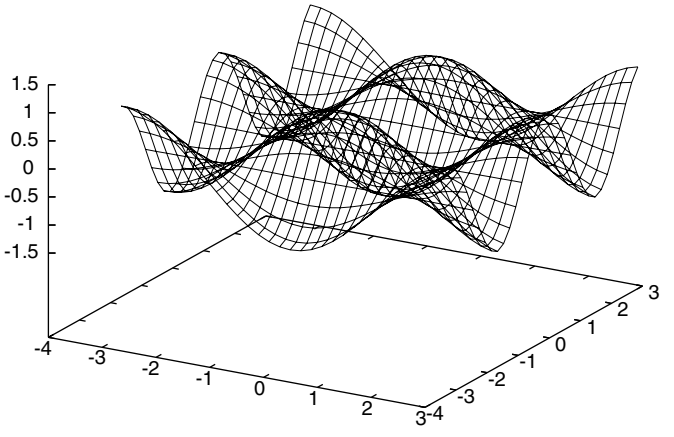
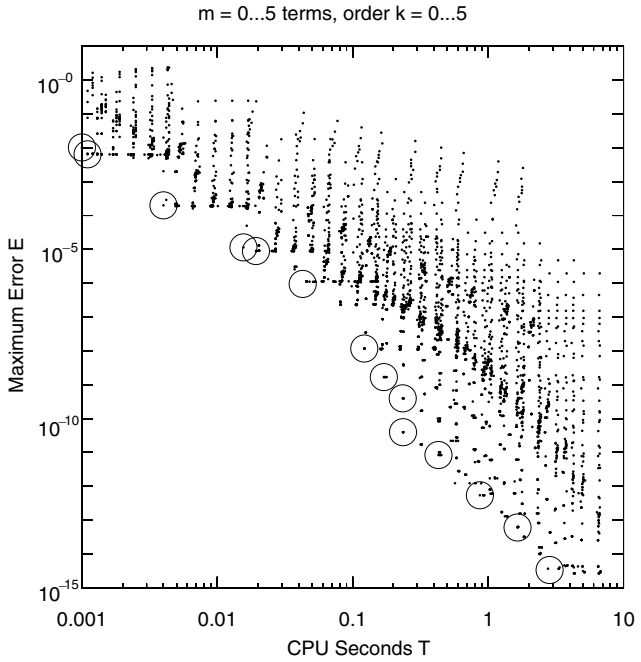


Fig. 1. A random C^4 solution u of the Poisson equation.



n	02	03	04	05	06	07	08	09	10	11	12	13	14
T	.0013	.0012	.0049	.0202	.0194	.0803	.122	.174	.234	.236	.44	.92	1.65
m	1	3	3	3	4	5	3	4	4	5	5	5	5
k	1	1	1	1	1	1	1	1	2	2	2	3	5
N	8	8	16	32	32	64	80	96	96	96	128	160	160
τ	.23-1	.78-3	.16-4	.49-4	.49-6	.12-4	.79-7	.54-7	.54-6	.54-7	.31-7	.20-7	.20-7

Fig. 2. A log–log plot of error E vs. CPU time T (marked by circles in the plot) for n-digit accuracy, for a random C^4 solution u of the Poisson equation (Fig. 1).

12-digit accuracy most efficiently. The accompanying table of optimal parameters shows the CPU time and parameter values which most efficiently achieve n-digit accuracy for $n = 2, 3, \ldots, 14$. The fast convergence to roundoff level exhibited by the solution and its derivatives, for fine meshes and high-order local correction, demonstrates the stability of our method.

11. Boundary integral equations and variable coefficients

Our locally-corrected spectral methods extend to the solution of linear second-order elliptic systems (1) with standard boundary conditions on arbitrary d-dimensional domains $\Omega \subset Q$. After conversion to first order, such problems include zero-order boundary conditions $B(\gamma)u(\gamma) = g(\gamma)$ on $\Gamma = \partial\Omega$. Without loss of generality, we can assume $B\,B^* = I$, and define local projections $Q = B^*B$ and $P = I - Q$ at each boundary point $\gamma \in \Gamma$. Then Eq. (5) becomes a boundary integral equation

$$\frac{1}{2}\mu(\gamma) + \int_\Gamma \sum_{j=1}^{d} n_j(\sigma)\varphi_\gamma(\sigma)A_j\mu(\sigma)\,\mathrm{d}\sigma = \rho(\gamma), \tag{14}$$

for the new projected unknown $\mu(\gamma) = P(\gamma)u(\gamma)$. The right-hand-side is a combination of volume and layer potentials

$$\rho(\gamma) = \int_D \varphi_\gamma(y)f(y)\,\mathrm{d}y - \frac{1}{2}B^*g(\gamma) - \int_\Gamma \sum_{j=1}^{d} n_j(\sigma)\varphi_\gamma(\sigma)A_jB^*g(\sigma)\,\mathrm{d}\sigma,$$

and μ also satisfies a local condition $B\mu = 0$ at each point of Γ. The factor 1/2 applies to smooth domains Ω. At corners of nonsmooth domains, it becomes the fraction of solid angle subtended by the corner.

Our boundary integral Eq. (14) simplifies classical potential theory by eliminating double layer potentials. In the special case of the time-harmonic Maxwell equations (Example 3), it resembles the combined field approach which resolves some well-known difficulties with resonances in computational electromagnetism [35].

We then apply our Ewald summation technique for φ to separate the kernel of the boundary integral equation (14),

$$K(\gamma, \sigma) = \sum_{j=1}^{d} n_j(\sigma)\varphi_\gamma(\sigma)A_j,$$

into a global rapidly-converging Fourier series and a local correction. The Fourier series is a low-rank kernel, while the local correction is a combination of δ-functions and their derivatives on the interface. In the Poisson and Stokes cases, our local correction generalizes other techniques for boundary integrals [36,28]. Thus our technique expresses the kernel as a global low-rank modification of a local differential operator, and permits the application of standard fast solution techniques.

Our new approach also extends to periodic variable-coefficient problems as in [37]: We represent the solution as a volume potential formed with the locally-corrected fundamental solution of a conveniently chosen constant-coefficient problem, varying on subdomains to capture local behavior of the coefficients. We solve a *locally* implicit volume potential equation by iteration or direct low-rank updating, and evaluate the potential as desired.

12. Conclusions

We have presented a new approach to the solution of general elliptic systems, based on conversion to overdetermined first-order systems and a new Ewald summation technique. Numerical results in periodic geometry demonstrate the accuracy and efficiency of this approach. We have also derived boundary integral equations which employ the new approach to solve general elliptic systems on complex domains.

Acknowledgments

We thank J.T. Beale, T. Chen, G. Papanicolaou, I. Sammis and B. Simeon for valuable discussions, and the anonymous referees for many helpful suggestions. This material is based upon work supported by the National Science Foundation under Grant Number DMS-0512963, and by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under Grant Number FA9550-05-1-0120. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

References

[1] D.L. Clements, Boundary Value Problems Governed by Second Order Elliptic Systems, Pitman, London, 1981.
[2] A. Brandt, S.R. Fulton, G.D. Taylor, Improved spectral multigrid methods for periodic elliptic problems, J. Comput. Phys. 58 (1985) 96–112.
[3] E. Braverman, M. Israeli, A. Averbuch, L. Vozovoi, A fast 3D Poisson solver of arbitrary order accuracy, J. Comput. Phys. 144 (1998) 109–136.
[4] J. Strain, A boundary integral approach to unstable solidification, J. Comput. Phys. 85 (1989) 342–389.
[5] J. Strain, A fast semi-Lagrangian contouring method for moving interfaces, J. Comput. Phys. 169 (2001) 1–22.
[6] A.W. Bargteil, T.G. Goktekin, J.F. O'Brien, J.A. Strain, A semi-Lagrangian contouring method for fluid simulation, ACM Trans. Graphics 25 (2006) 1–22.
[7] G.N. Hile, M.H. Protter, Properties of overdetermined elliptic systems, Arch. Ration. Mech. Anal. 66 (1977) 267–293.
[8] M.H. Protter, Overdetermined first order elliptic systems, in: P.W. Schaefer (Ed.), Maximum Principles and Eigenvalue Problems in Partial Differential Equations, Longman Scientific and Technical, Harlow, Essex, 1988, pp. 68–81.
[9] C. Cosner, On the definition of ellipticity for systems of partial differential equations, J. Math. Anal. Appl. 158 (1991) 80–93.
[10] K. Krupchyk, W. Seiler, J. Tuomela, Overdetermined elliptic systems, Found. Comp. Math. (to appear).

[11] J.J. Heys, T. Manteuffel, S. McCormick, J. Ruge, First-order system least squares for coupled fluid-elasticity problems, J. Comput. Phys. 195 (2004) 560–575.
[12] M. Costabel, M. Dauge, Crack singularities for general elliptic systems, Math. Nach. 235 (2002) 29–49.
[13] F. John, Partial Differential Equations, 4th ed., Springer Verlag, New York, 2005.
[14] J. Bazer, R. Burridge, Energy partition in the reflection and refraction of plane waves, SIAM J. Appl. Math. 34 (1978) 78–92.
[15] K. Sandberg, Forward and inverse wave propagation using bandlimited functions and a fast reconstruction algorithm for electron microscopy, Ph.D. thesis, University of Colorado, 2003.
[16] T. Kato, Perturbation Theory for Linear Operators, Springer, 1976.
[17] A. Ben-Israel, T.N.E. Greville, Generalized Inverses, Springer Verlag, 2003.
[18] C.-D. Munz, F. Kemm, R. Schneider, E. Sonnendrucker, Divergence corrections in the numerical simulation of electromagnetic wave propagation, in: Hyperbolic Problems: Theory, Numerics, ApplicationsProceedings of the 8th International Conference on Hyperbolic Problems, Birkhauser Verlag, 2001, pp. 603–612.
[19] B. Fornberg, A Practical Guide to Pseudospectral Methods, Cambridge, 1996.
[20] J.P. Boyd, Chebyshev and Fourier Spectral Methods, Dover, 2001.
[21] P. Ewald, Die Berechnung optischer und elektrostatistischer Gitterpotentiale, Ann. Phys. 64 (1921) 253.
[22] D.J. Adams, G.S. Dubey, Taming the Ewald sum in the computer simulation of charged systems, J. Comput. Phys. 72 (1987) 156–176.
[23] C. Holm, Efficient methods for long range interactions in periodic geometries plus one application, in: Computational Soft Matter: From Synthetic Polymers to ProteinsNIC Lecture Notes, vol. 23, John von Neumann Institute for Computing, Julich, 2004, pp. 195–236.
[24] H. Hasimoto, On the periodic fundamental solutions of the Stokes equations and their applications to viscous flow past a cubic array of spheres, J. Fluid Mech. 5 (1959) 317–326.
[25] D. Saintillan, E. Darve, E.S.G. Shaqfeh, A smooth particle-mesh Ewald algorithm for Stokes suspension flows: The sedimentation of fibers, Phys. Fluids 17 (2004) 1–21.
[26] J. Strain, Fast potential theory I: Poisson solvers on a cube, Report PAM-480, Center for Pure and Applied Mathematics, UC Berkeley, December 1989.
[27] J. Strain, Fast potential theory II: layer potentials and discrete sums, J. Comput. Phys. 99 (1992) 251–270.
[28] J.T. Beale, J. Strain, Locally-corrected semi-Lagrangian methods for Stokes flow with elastic interfaces, (in preparation).
[29] J. Stoer, R. Bulirsch, Introduction to Numerical Analysis, Springer-Verlag, 1993.
[30] I.S. Sammis, J. Strain, Piecewise-polynomial non-uniform fast fourier transforms, (in preparation).
[31] R.B. Sidje, Expokit: software package for computing matrix exponentials, ACM Trans. Math. Software 24 (1998) 130–156.
[32] G.H. Golub, C.F. van Loan, Matrix Computations, 2nd ed., Johns Hopkins University Press, Baltimore, 1989.
[33] J. Strain, Locally-corrected multidimensional quadrature rules for singular functions, SIAM J. Sci. Comput. 16 (1995) 992–1017.
[34] B. Fornberg, Generation of finite difference formulas on arbitrarily spaced grids, Math. Comput. 51 (1988) 699–706.
[35] B. Beker, K.R. Umashankar, A. Taflove, Numerical analysis and validation of the combined field surface integral equations for electromagnetic scattering by arbitrary shaped two-dimensional anisotropic objects, IEEE Trans. Ant. Prop. 37 (1989) 1573–1582.
[36] J.T. Beale, M.C. Lai, A method for computing nearly singular integrals, SIAM J. Numer. Anal. 38 (2001) 1902–1925.
[37] J. Strain, Fast spectrally-accurate methods for variable-coefficient elliptic problems, Proc. Am. Math. Soc. 122 (1995) 843–850.