Results in Applied Mathematics 15 (2022) 100288

Contents lists available at ScienceDirect

Results in Applied Mathematics

journal homepage: www.elsevier.com/locate/results-in-applied-mathematics

On the fundamental solutions-based inversion of Laplace matrices

F.J. Vermolen [a,b,*], D.R. den Bakker [b], C. Vuik [b]

[a] University of Hasselt, Department of Mathematics and Statistics, Diepenbeek, Belgium
[b] Delft Institute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Article history:
Received 7 February 2022
Received in revised form 13 May 2022
Accepted 16 May 2022
Available online xxxx

Keywords:
Inverse matrix
Laplace matrix
Finite element discretisation
Fundamental solutions

ABSTRACT

The discretisation of the Laplacian results into the well-known Laplace matrix. In the case of a one dimensional problem, an explicit formula for its inverse is derived on the basis of fundamental solutions (Green's functions) for general boundary conditions. For a linear reaction–diffusion equation, approximations of the inverse are given.

1. Introduction

Many partial differential equations contain the Laplace operator. Such differential equations model (diffusive) transport of matter, energy (heat) or momentum. In the computation of approximations for solutions of these equations, one often has to solve (large) systems of linear algebraic equations in order to get the desired quantity on the mesh points. These equations are often characterised by the discretised Laplacian, which is also commonly referred to as the Laplace matrix. Solution procedures often take advantage of the symmetry and positive definiteness of the Laplace matrix, which results from the self-adjointness and positive definiteness of the (continuous) Laplace operator. Well-known solution procedures are direct methods such as Gaussian elimination, Thomas algorithm (for a tridiagonal system) and Choleski decomposition, see for instance [1–3]. If the system is very large, then one uses classical iterative methods such as defect correction, Gauss, Gauss–Seidel, SOR, (preconditioned) Krylov subspace methods (such as conjugate gradients), or multigrid methods, see for instance [3–6], among very many others. One of the major issues in solving these large systems of equations is that the discretisation matrix becomes ill-conditioned as the resolution increases, which makes it necessary to perform a large number of iterations or to apply preconditioning in order to reach a predefined accuracy. In order to speed up computations, often one uses preconditioning, in which the algebraic system of equations, given by

$$S\mathbf{x} = \mathbf{b}, \tag{1}$$

is multiplied by a matrix P that approximates the inverse of S. The idea is to improve the effective condition of the resulting system of equations by decreasing the width (that is the ratio between the largest and smallest eigenvalues) of the spectrum of the eigenvalues. For preconditioning, one uses approximations like Incomplete Choleski, diagonal preconditioning, Eissenstadt, or even deflation methods to 'remove' the last, few persisting smallest eigenvalues, see for instance [7,8]. The eigenspace from the smallest, persistent eigenvalues is typically approximated so that the eigenvalues

* Corresponding author at: University of Hasselt, Department of Mathematics and Statistics, Diepenbeek, Belgium.
E-mail address: Fred.Vermolen@uhasselt.be (F.J. Vermolen).

https://doi.org/10.1016/j.rinam.2022.100288

and eigenvectors from the smallest eigenvalues do not have to be determined. Deflation is often applied in physical systems with very sharp transitions, such as in porous media, see for instance [9].

The current communication focuses on analytic expressions for approximations of the inverse of Laplace-like matrices. We do not claim to be the first authors that present a closed form expression of the inverse of Laplace-like matrices, however, we believe that our approach is simple and elegant, and straightforward to carry out. For instance, Gueye [10] found an exact representation of the inverse of the matrix that was obtained after finite difference discretisation of the 1D Laplace/Poisson equation by means of Gaussian elimination and substitution. Our current approach is based on fundamental solutions (Green's functions) and allows more general configurations. Furthermore, cases with combinations with generic linear boundary conditions (Dirichlet, Neumann, Robin) can be dealt with, as well as non-uniform finite element meshes. In the case of one-dimensional problems, an exact representation for the inverse of the Laplace matrix is derived. The derivation is based on Green's functions of the (continuous) differential equation, see for instance [11]. The inverse enables us to represent solutions to the Laplace matrix in an algebraic way, which is an alternative next to more classical ways. Approximations, as well as their accuracy, are derived for (linear) reaction–diffusion equations. The obtained formulas for the inverses of matrices may be used in mathematical proofs that need the inverse of Laplace matrices. An example of such an application is where one derives conditions for monotonicity of solutions of linear systems that result from the (finite element or (staggered) finite volume) discretisation of saddle point problems. This issue is relevant for the numerical solution of Biot's poroelasticity model, where stabilisation techniques have to be used to warrant monotonic finite element solutions, see [12–14] for instance.

First we will start with the procedure, in which we will derive the Green's matrices as an algebraic analogue of Green's functions (fundamental solutions) in the context of differential equations. Then, we will outline the procedure for linear reaction–diffusion equations and study convergence both analytically and experimentally.

2. The general principle

The general idea behind our approach is the following. We consider a (finite element) discretisation method to approximate the solution of a (partial) differential equation, which results into a mesh with nodal points in the domain of computation. We assume that the number of unknowns is given by n. Suppose that a matrix $S \in \mathbb{R}^{n\times n}$ is given. This matrix S is the discrete counterpart of a linear differential operator, L, including (homogeneous) boundary conditions. Let the solution be known for a given right-hand side, in other words, let the function u_i be known for a given right-hand side f_i:

$$Lu_i = f_i, \text{ in } \Omega,$$

where Ω represents the domain of computation. We assume that the above problem has a unique solution (that is $Lu = 0 \iff u = 0$ in Ω). Let S denote the non-singular finite element discretisation matrix of the above problem, and let $\mathbf{b}_i$ denote the right-hand side that corresponds to the above continuous problem, then we have

$$S\mathbf{v}_i = \mathbf{b}_i,$$

Imagine further that the solution u_i is such that, on the finite element nodal points, it is equal to the solution of the discretised problem, that is $u_i(\vec{x}_j) = v_{ij}$, where $\mathbf{v}_i = [v_{i1}\ v_{i2}\ \dots\ v_{in}]^T$. The idea is that we use n equations in which all the f_i and $\mathbf{b}_i$ are linearly independent. This gives n linearly independent solutions for $\mathbf{v}_i$ and u_i. Putting these expressions in matrices, that is

$$B = [\mathbf{b}_1 \dots \mathbf{b}_n], \text{ and } V = [\mathbf{v}_1 \dots \mathbf{v}_n],$$

gives

$$SV = B \iff S^{-1} = VB^{-1}.$$

If B^{-1} is easier to invert than S, which is true if B is diagonal or equal to the identity matrix, then this operation pays off, else, this operation is unpractical. In the setting of Green's functions, we will use finite elements and obtain an identity matrix for B. The idea works perfectly if the solution to the discretised problem is the same as the solution of the continuous problem. First, we will work this out for the one-dimensional Laplacian operator.

3. The exact representation of 1D Laplace matrices

The exact representation formulas for 1D Laplace under generic boundary conditions and under generic meshes are shown. In a one-dimensional setting, we take $\Omega = (0, 1)$, with closure $\overline{\Omega} = [0, 1]$, and we consider the following Poisson equation

$$\begin{cases} -u'' = \delta(x - s), \quad x \in \Omega, \quad s \in \overline{\Omega}, \\ \end{cases} \tag{2}$$

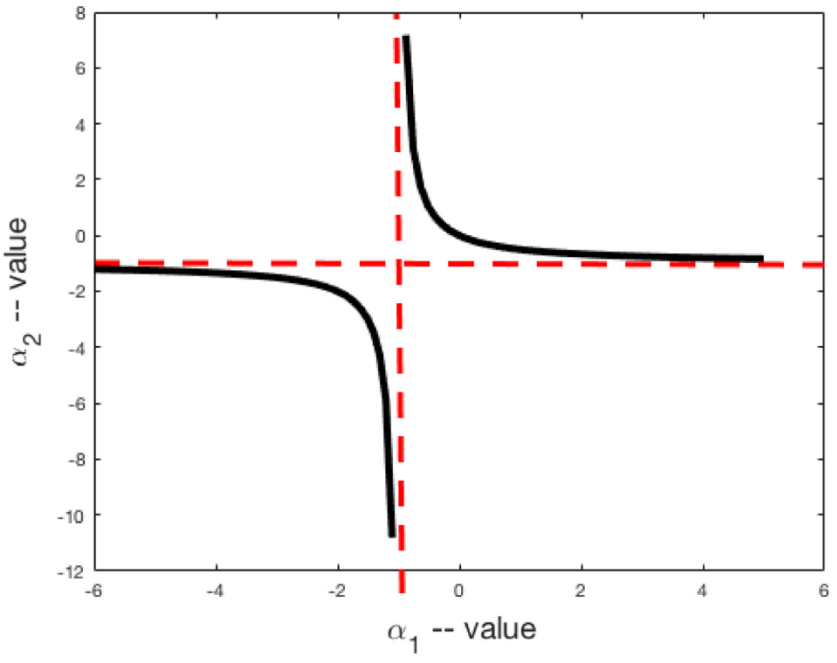

Fig. 1. The 'forbidden region', defined by $\alpha_1+\alpha_2+\alpha_1\alpha_2=0$, indicated by the black, solid curve. The asymptotes $\alpha_1=-1$ and $\alpha_2=-1$ are indicated by the red, dashed curves.

Here the prime stands for differentiation with respect to x, and $\delta(.)$ denotes the Dirac delta (probability) distribution, which is characterised by

$$\begin{cases} \delta(x)=0, \text{ for all } x\neq 0 \text{ and for each open domain, } \Omega, \text{ containing } x=0, \\ \text{we have } \int_\Omega \delta(x)\,dx=1. \end{cases} \tag{3}$$

Let

$$\mathbb{V} := \{(\alpha_1,\alpha_2)\in\mathbb{R}^2 : \alpha_1,\alpha_2\geq 0 \text{ and } \alpha_1+\alpha_2>0\},$$

then it follows that if $(\alpha_1,\alpha_2)\in\mathbb{V}$ then the differential operator in Problem (2) is coercive (positive definite), and hence Problem (2) has a uniquely defined solution. The region $\mathbb{V}$ is referred to as 'the region of positive definiteness'. The exact solution to boundary value problem (2) is given by

$$u(x;s)=\frac{1+\alpha_2(1-s)}{\alpha_1+\alpha_2+\alpha_1\alpha_2}(\alpha_1 x+1)-(x-s)_+, \tag{4}$$

provided that $\alpha_1+\alpha_2+\alpha_1\alpha_2\neq 0$. The 'forbidden region', $\mathbb{F}$, defined by

$$\mathbb{F} := \{(\alpha_1,\alpha_2)\in\mathbb{R}^2 : \alpha_1+\alpha_2+\alpha_1\alpha_2=0\},$$

has been plotted in Fig. 1. In the above equation, we used the convention $(.)_+ := \max(0,.)$. This solution can be obtained easily by integration twice or by the use of Laplace transformations. Sending α_1 and α_2 to infinity provides us with boundary condition $u(0)=0$ and $u(1)=1$, respectively. For the case of Dirichlet boundary conditions on both boundary points, that is both α_1 and α_2 are sent to infinity, we have

$$u(x;s)=(1-s)x-(x-s)_+. \tag{5}$$

In order to keep the problem positive definite (coercive), one requires $(\alpha_1,\ \alpha_2)\in\mathbb{V}$, which is a sufficient (hence not necessary) condition for the existence of a uniquely defined solution. If $(\alpha_1,\ \alpha_2)\in\mathbb{V}$, then $\alpha_1+\alpha_2+\alpha_1\alpha_2\geq\alpha_1+\alpha_2>0$, and hence it is clear that the 'forbidden region' in Fig. 1 has no overlap with the 'region of positive definiteness', $\mathbb{V}$. This implies that

$$(\alpha_1,\alpha_2)\in\mathbb{V}\Longrightarrow(\alpha_1,\alpha_2)\notin\mathbb{F} \text{ (in particular } \alpha_1+\alpha_2+\alpha_1\alpha_2>0).$$

From Eq. (4), it is clear that the solution exhibits a piecewise linear behaviour. This means that for $x\neq s$, all derivatives ... the weak solution, the solution is sought in function spaces, which for the benefit

of the reader, we define below. Suppose that $\Omega \subset \mathbb{R}^d$ is a open, bounded Lipschitz domain in the d–dimensional space, then the space of L^p–integrable functions is defined by:

$$L^p(\Omega) := \{f : \Omega \longrightarrow \mathbb{R} \ : \ \int_\Omega |f|^p d\Omega < \infty\}.$$

The finite element space, where we seek the solution, is defined by

$$H^1(\Omega) = \{f \in L^2(\Omega) \ : \ \frac{\partial f}{\partial x_i} \in L^2(\Omega), \ \forall i \in \{1, \ldots, d\}\}.$$

Then, with $\Omega = (0, 1)$, the weak form of problem (2) is given by

$$\left\{ \begin{array}{l} \text{Find } u \in H^1(\Omega) \text{ such that} \\ \alpha_1 u(0)\phi(0) + \alpha_2 u(1)\phi(1) + \int_0^1 u'\phi' dx = \delta_s(\phi) = \phi(s), \\ \quad \forall \phi \in H^1(\Omega), \quad s \in [0, 1]. \end{array} \right. \tag{6}$$

Let $\Omega \subset \mathbb{R}^d$ be open, bounded and Lipschitz in $\mathbb{R}^d$, then the Sobolev space $W^{k,p}(\Omega)$ is defined by

$$W^{k,p}(\Omega) := \{f \in L^p(\Omega) \ : \ D^\alpha f \in L^p(\Omega), \ \forall \alpha \leq k\},$$

where α denotes the multi-index, $\alpha = (\alpha_1, \ldots, \alpha_d)$ (not to be confused with the parameters in the boundary conditions), $|\alpha| = \sum_{j=1}^d \alpha_j$ and $D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}$. Further, the infinity norm is defined by

$$L^\infty(\Omega) = \inf_{C \geq 0} \{ \ |f| \leq C, \quad \text{a.e. in} \quad \Omega \ \}.$$

Sobolev's Inequality (see for instance Brenner & Scott [15], page 33) states that there is a $C > 0$ such that

$$\forall \phi \in W^{k,p}(\Omega) : \qquad \|\phi\|_{L^\infty(\Omega)} \leq C \|\phi\|_{W^{k,p}(\Omega)}, \qquad \text{if } d < k\, p,$$

where d, k and p, respectively, denote the dimensionality (which is 1 here), the order of the derivative and the exponent of the Hölder norm. Since we have $H^1(\Omega) = W^{1,2}(\Omega)$, hence $k = 1$, $p = 2$ and $d = 1$, it follows that the above Sobolev's Inequality is satisfied, and hence

$$|\phi(s)| \leq \|\phi\|_{L^\infty(\Omega)} \leq C \|\phi\|_{H^1(\Omega)},$$

which says that the right-hand side functional $\delta_s(\phi)$ is bounded in $H^1(\Omega)$. We also remark that if $(\alpha_1, \alpha_2) \in \mathbb{V}$, then Lax–Milgram's Theorem provides the existence and uniqueness of a solution in $H^1(\Omega)$, if $\Omega \subset \mathbb{R}$. Note that for dimensionalities higher than one, boundedness in $H^1(\Omega)$ does not follow. For the time being we consider $\Omega = (0, 1)$, for the case that $\phi \in H^1(\Omega)$ under the restriction that $\phi(0) = 0$, it follows that

$$|\phi(s)| = |\int_0^s \phi'(t) dt| \leq \sqrt{s} \left[\int_0^s (\phi'(t))^2 dt \right]^{1/2} \leq \left[\int_0^1 (\phi'(t))^2 dt \right]^{1/2} \leq \|\phi\|_{H^1(0,1)},$$

and hence $C = 1$ in this case. The idea is that we assemble a finite element mesh with n linear Lagrangian elements, in which s represents any nodal point of the finite element mesh, for which we have $x_{k-1} < x_k$, $x_1 = 0$, $x_n = 1$ and $h_i = x_{i+1} - x_i$ for $i = 1, \ldots, n-1$. By locating the force of the delta distribution on the mesh points, that is $s = x_j$, we construct the following solutions using Eq. (4):

$$u(x_i; x_j) = \frac{1 + \alpha_2(1 - x_j)}{\alpha_1 + \alpha_2 + \alpha_1\alpha_2}(\alpha_1 x_i + 1) - (x_i - x_j)_+, \tag{7}$$

The Galerkin method, where the numerical approximation, u^h, is sought as a linear combination of the linear Lagrangian basis functions, dictates that $u^h(x) = \sum_{j=1}^n c_j \phi_j(x)$. This implies that the algebraic system for c_j is given by

$$S\mathbf{c} = \mathbf{b}, \tag{8}$$

where S_{ij} is characterised as follows:

$$S_{11} = \alpha_1 + \frac{1}{h_1}, \ S_{ii} = \frac{1}{h_{i-1}} + \frac{1}{h_i}, \ \text{for } i = 2, \ldots, n-1,$$

$$S_{nn} = \alpha_2 + \frac{1}{h_{n-1}}, \ S_{i-1i} = S_{ii-1} = \frac{1}{h_{i-1}}, \ \text{for } i = 2, \ldots, n,$$

$$S_{ij} = 0, \ \text{for } |i - j| > 1.$$

The right-hand side vector $\mathbf{b}$ is characterised by

where $\mathbf{e}_i$ is the unit vector in the ith direction, that is, all components of $\mathbf{e}_i$ are zero, except for the ith component, which is equal to one. Choosing the delta distributions to act on the finite element mesh points, and using Céa's Lemma, the H^1–error of the interpolatory solution, Aubin–Nitsche's trick, one arrives at a quadratic error in the L^2–norm of the finite element solution [16], taking $\Omega = (0, 1)$:

$$0 \leq \|u - u^h\|_{L^2(\Omega)} \leq Kh^2 \sum_{k=1}^{n-1} \|u''\|_{L^2(\Omega_k)}, \tag{9}$$

where $\Omega_k = (x_k, x_{k+1})$. Further, h is the maximum element size, and since the second derivative vanishes in each element, the error vanishes and therefore the finite-element solution is equal to the (exact) solution (4). It is noted that it is crucially important to choose the point of action of the Dirac pulse on the nodal points, otherwise $u \notin H^2(\Omega_k)$. Collecting the equations for all $i = 1, \ldots, n$, gives the identity matrix on the right-hand side. Furthermore since the error vanishes, it immediately follows that the columns of the inverse of the discretisation matrix S are given by the exact solution on the nodal points of the finite element mesh. Hence, using Eq. (7), upon substituting $s = x_j$, we get

$$(S^{-1})_{ij} = u(x_i; x_j) = \frac{1 + (1 - x_j)\alpha_2}{\alpha_1 + \alpha_2 + \alpha_1\alpha_2}(\alpha_1 x_i + 1) - (x_i - x_j)_+, \qquad i, j = 1, \ldots, n, \tag{10}$$

where $x_i = \sum_{k=1}^{i-1} h_k$, which implies the following theorem:

Theorem 3.1. *Let $S \in \mathbb{R}^{n\times n}$, $h_i > 0$ for $i = 1, \ldots, n-1$, and let S be given by*

$$S = \begin{pmatrix} \alpha_1 + \frac{1}{h_1} & -\frac{1}{h_1} & 0 & \ldots & \ldots & 0 \\ -\frac{1}{h_1} & \frac{1}{h_1} + \frac{1}{h_2} & -\frac{1}{h_2} & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & -\frac{1}{h_{n-2}} & \frac{1}{h_{n-2}} + \frac{1}{h_{n-1}} & -\frac{1}{h_{n-1}} \\ 0 & \ldots & \ldots & 0 & -\frac{1}{h_{n-1}} & \frac{1}{h_{n-1}} + \alpha_2 \end{pmatrix},$$

then its inverse exists, and the inverse is given by

$$(S^{-1})_{ij} = \frac{1 + (1 - \sum_{k=1}^{j-1} h_k)\alpha_2}{\alpha_1 + \alpha_2 + \alpha_1\alpha_2}(\alpha_1 \sum_{k=1}^{i-1} h_k + 1) - (\sum_{k=1}^{i-1} h_k - \sum_{k=1}^{j-1} h_k)_+, \qquad i, j = 1, \ldots, n,$$

provided that $\alpha_1 + \alpha_2 + \alpha_1\alpha_2 \neq 0$.

Note that S^{-1} is a full (non-sparse) matrix. For an equidistant mesh (with $(n-1)h = 1$), one obtains

$$(S^{-1})_{ij} = \frac{1 + (1 - (j-1)h)\alpha_2}{\alpha_1 + \alpha_2 + \alpha_1\alpha_2}(\alpha_1(i-1)h + 1) - h(i-j)_+, \qquad i, j = 1, \ldots, n,$$

Sending α_1 and α_2 to infinity, which entails homogeneous Dirichlet boundary conditions at both sides. For the sake of simplicity, we will count the unknowns differently by $x_0 = 0 < x_1 < x_2 < \cdots < x_{n-1} < x_n < x_{n+1} = 1$ such that unknowns are only positioned on $x_1, \ldots, x_n$. Using $h_i = x_{i+1} - x_i$ for $i = 0, \ldots, n$, we get

$$S = \begin{pmatrix} \frac{1}{h_0} + \frac{1}{h_1} & -\frac{1}{h_1} & 0 & \ldots & \ldots & 0 \\ -\frac{1}{h_1} & \frac{1}{h_1} + \frac{1}{h_2} & -\frac{1}{h_2} & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & -\frac{1}{h_{n-2}} & \frac{1}{h_{n-2}} + \frac{1}{h_{n-1}} & -\frac{1}{h_{n-1}} \\ 0 & \ldots & \ldots & 0 & -\frac{1}{h_{n-1}} & \frac{1}{h_{n-1}} + \frac{1}{h_n} \end{pmatrix},$$

Using Eq. (5), substituting x_j and x_i for s and x, respectively, and using $x_i = \sum_{k=0}^{i-1} h_k$, the inverse of S is given by

$$(S^{-1})_{i,j} = (1 - \sum_{k=0}^{j-1} h_k) \sum_{k=0}^{i-1} h_k - (\sum_{k=0}^{i-1} h_k - \sum_{k=0}^{j-1} h_k)_+, \qquad i, j = 1, \ldots, n,$$

The case where $h_i = h$ (and $(n+1)h = 1$), also follows easily:

$$(S^{-1})_{i,j} = (1 - jh)ih - h(i-j)_+, \qquad i, j = 1, \ldots, n.$$

Regarding the traditional Laplace matrix, which is just the same as the double Dirichlet case with an equidistant mesh, times h,

$$A = \begin{pmatrix} 2 & -1 & 0 & 0 & \ldots & 0 \\ -1 & 2 & -1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ & & & & & -1 \\ & & & & & 2 \end{pmatrix} = hS. \tag{11}$$

This gives, with using $(n+1)h = 1$, the following inverse

$$(A^{-1})_{ij} = \frac{n+1-j}{n+1} i - (i-j)_+.$$

We remark that this procedure is based on a Lagrangian-based finite element representation of the problem, which allows the very straightforward treatment of the Dirac delta distribution. The above expression was also obtained in [10], and applies to finite element methods and apart from a factor h, it also can also be applied to finite differences. The (exact) solution 'lives' in the finite element space $H^1(\Omega)$ in the 1D case. In the case of higher dimensionality, this is not possible since the fundamental solution does not live in $H^1(\Omega)$ and has a singularity in the point of action of the Dirac delta distribution, which makes that Sobolev's Inequality is not satisfied.

4. The approximate inverse for Laplace-reaction problems and error analysis

First we show the formulas for the approximate inverse. This is followed by an error analysis.

4.1. The approximate inverse

We consider the following problem on $\Omega = (0, 1)$:

$$\begin{cases} -u'' + \beta\, u = \delta(x-s), \text{ with } 0 \le s \le 1, \\ -u'(0) + \alpha_1 u(0) = 0, \\ u'(1) + \alpha_2 u(1) = 0. \end{cases} \tag{12}$$

We take $\beta > 0$. Processing the boundary conditions, gives the following general solution

$$u(x;s) = \frac{u(0;s)}{\sqrt{\beta}} \sinh(\sqrt{\beta}x) + \alpha_1 u(0;s)\cosh(\sqrt{\beta}x) - \frac{1}{\sqrt{\beta}} \sinh(\sqrt{\beta}(x-s)) \cdot H(x-s), \tag{13}$$

where

$$u(0;s) = \frac{\cosh(\sqrt{\beta}(1-s)) + \frac{\alpha_2}{\sqrt{\beta}} \sinh(\sqrt{\beta}(1-s))}{(1+\alpha_1\alpha_2)\cosh(\sqrt{\beta}) + (\alpha_1\sqrt{\beta} + \frac{\alpha_2}{\sqrt{\beta}})\sinh(\sqrt{\beta})}. \tag{14}$$

The solution shows that all its derivatives of orders that are larger or equal to two with respect to x do not vanish and hence the finite element error will not be zero. Herewith the finite element solution is not equal to the solution of the boundary value problem. The idea is to approximate the inverse of the discretisation matrix by the use of Green's functions. The weak form is given by

$$\begin{cases} \text{Find } u \in H^1(\Omega) \text{ such that} \\ \alpha_1 u(0)\phi(0) + \alpha_2 u(1)\phi(1) + \displaystyle\int_0^1 u'\phi' + \beta u\phi dx = \delta_s(\phi) = \phi(s), \\ \quad \forall \phi \in H^1(\Omega), \quad s \in [0,1]. \end{cases} \tag{15}$$

Using linear Lagrangian element basis functions, gives the following discretisation matrix for S

$$S_{11} = \alpha_1 + \frac{1}{h_1} + \beta\frac{h_1}{3},\ S_{ii} = \frac{1}{h_{i-1}} + \frac{1}{h_i} + \frac{\beta}{3}(h_{i-1} + h_i),\ \text{for } i = 2, \ldots, n-1,$$

$$S_{nn} = \alpha_2 + \frac{1}{h_{n-1}} + \frac{\beta}{3}h_{n-1},\ S_{i-1i} = S_{ii-1} = \frac{1}{h_{i-1}} + \frac{\beta}{6}h_{i-1},\ \text{for } i = 2, \ldots, n,$$

$$S_{ij} = 0,\ \text{for } |i-j| > 1.$$

By locating the force of the delta distribution on the mesh points, that is $s = x_j$, we construct the following solutions using Eqs. (13) and (14):

$$u(x_i;x_j) = \frac{u(0;x_j)}{\sqrt{\beta}} \sinh(\sqrt{\beta}x_i) + \alpha_1 u(0;x_j)\cosh(\sqrt{\beta}x_i) - \frac{1}{\sqrt{\beta}} \sinh(\sqrt{\beta}(x_i - x_j)) \cdot H(x_i - x_j), \tag{16}$$

where

$$\frac{\cosh(\sqrt{\beta}(1-x_j)) + \frac{\alpha_2}{\sqrt{\beta}} \sinh(\sqrt{\beta}(1-x_j))}{(\alpha_1\sqrt{\beta} + \frac{\alpha_2}{\sqrt{\beta}})\sinh(\sqrt{\beta})}. \tag{17}$$

It is straightforward to show that Eqs. (16) and (17) converge to Eq. (7) as $\beta \longrightarrow 0$. Once again, the idea is to approximate the inverse of S by the above solution

$$(S^{-1})_{ij} \approx \frac{u(0;x_j)}{\sqrt{\beta}} \sinh(\sqrt{\beta}x_i) + \alpha_1 u(0;x_j)\cosh(\sqrt{\beta}x_i) - \frac{1}{\sqrt{\beta}} \sinh(\sqrt{\beta}(x_i - x_j)) \cdot H(x_i - x_j) = (\tilde{S}^{-1})_{ij}. \tag{18}$$

Here $\tilde{S}^{-1}$ is referred to as the approximate inverse of S. In the next section, we will derive an upper bound for the error of this approximation of the inverse of the matrix (that is, the difference between the inverse and the approximate inverse).

4.2. Error estimation

The Trapezoidal Rule is a Newton–Cotes quadrature rule that is based on linear interpolation using the mesh points x_i and x_{i+1}. To estimate the integration (quadrature) error, the interpolation error is used. Let u and u_h, respectively, denote the true function and the interpolated approximation of u, then the interpolation error is given by $u(x) - u_h(x) = -\frac{1}{2}(x - x_i)(x - x_{i+1})u''(\xi)$ for some $\xi \in (x_i, x_{i+1})$ (see for instance [3]). Integration of the result over the interval (x_i, x_{i+1}) gives $E(u) = -\frac{h^3}{12}u''(\xi)$ and hence, it follows that $|E(u)| \leq \gamma h^3$ for some $\gamma > 0$. Let u and u^h, now respectively, denote the solution and its finite element approximation and h denotes the mesh size, and let $w(x) = u(x) - u^h(x)$, then, using the Trapezoidal quadrature (Newton–Cotes) Rule and its accuracy of $\mathcal{O}(h^3)$ over a single element, Cauchy–Schwarz' Inequality and relation (9), one obtains

$$(EA)\begin{cases} |u(x_i) - u^h(x_i)| = |w(x_i)| \leq |w(x_i)| + |w(x_{i+1})| = \dfrac{2}{h} \cdot \dfrac{h}{2}(|w(x_i)| + |w(x_{i+1})|) = \\ \dfrac{2}{h}\left(\displaystyle\int_{x_i}^{x_{i+1}} |w(s)|ds + \gamma h^3\right) \leq \dfrac{2}{h}(\sqrt{h}\|w\|_{L^2(0,1)} + \gamma h^3) \leq \dfrac{2}{h}(\sqrt{h}Kh^2 + \gamma h^3) = \mathcal{O}(h^{3/2}). \end{cases}$$

In the literature, one can also find pointwise (or max norm) estimates of $\mathcal{O}(h^2 \log(\frac{1}{h}))$, see [17] with an extension to parabolic equations.

Regarding the numerical solution, we have the following

$$S\mathbf{u}_i^h = \mathbf{e}_i \text{ and } \tilde{S}\tilde{\mathbf{u}}_i = \mathbf{e}_i, \text{ where } \tilde{S} = (\tilde{S}^{-1})^{-1}. \tag{19}$$

Here u_i^h and $\tilde{u}_i$, respectively, represent the finite element solution and the solution to the differential equation with the Dirac delta distribution. From the definition of $\mathbf{e}_i$, it immediately follows that

$$\mathbf{u}_i^h = \begin{pmatrix} (S^{-1})_{1,i} \\ (S^{-1})_{2,i} \\ \cdots \\ (S^{-1})_{n,i} \end{pmatrix}, \text{ and } \tilde{\mathbf{u}}_i = \begin{pmatrix} (\tilde{S}^{-1})_{1,i} \\ (\tilde{S}^{-1})_{2,i} \\ \cdots \\ (\tilde{S}^{-1})_{n,i} \end{pmatrix} \tag{20}$$

Hence we have

$$(S^{-1})_{ji} = (\mathbf{u}_i^h)_j = u_i^h(x_j), \text{ and } (\tilde{S}^{-1})_{ji} = (\tilde{\mathbf{u}}_i^h)_j = \tilde{u}_i^h(x_j). \tag{21}$$

Using the pointwise error approximation (EA), gives

$$|(\tilde{S}^{-1})_{ij} - (S^{-1})_{ij}| = |\tilde{u}_i^h(x_j) - u_i^h(x_j)| = \mathcal{O}(h^{3/2}), \tag{22}$$

where we exploited symmetry of S^{-1} and $\tilde{S}^{-1}$, and hence the approximate inverse converges to the real inverse as $h \longrightarrow 0$, which is summarised in the following theorem:

Theorem 4.1. *Consider a linear Lagrangian finite element method for the boundary value problem (12) with $\beta > 0$, then the approximate inverse of the discretisation matrix $\tilde{S}^{-1}$ converges to the inverse S^{-1} as the mesh size h tends to zero, that is*

$$\lim_{h \longrightarrow 0} (\tilde{S}^{-1})_{ij} = (S^{-1})_{ij},$$

where the approximate inverse is defined in Eq. (18). Further, we have

$$|(\tilde{S}^{-1})_{ij} - (S^{-1})_{ij}| = \mathcal{O}(h^{3/2})$$

We remark that sharper bounds for the max norm of the finite element error based on linear Lagrangian elements in a domain in $\mathbb{R}^2$ can be found in, among others, [18], which reads as:

For all $\varepsilon > 0$, there is a $C_\varepsilon > 0$, such that $\|u - u_h\|_\infty \leq C_\varepsilon h^{2-\varepsilon}\|f\|_{H^2(0,1)}$, where f is the right-hand function. Since the norm on the right-hand side does not exist for the Dirac delta distribution, this sharp bound is not used. Furthermore, the matrix columns of the inverse of S represent the finite element solutions upon setting the Dirac pulses on the finite element mesh points, hence this bound also shows that the finite element errors are of order $\mathcal{O}(h^{3/2})$ if the actions of the Dirac pulses are not on the mesh points.

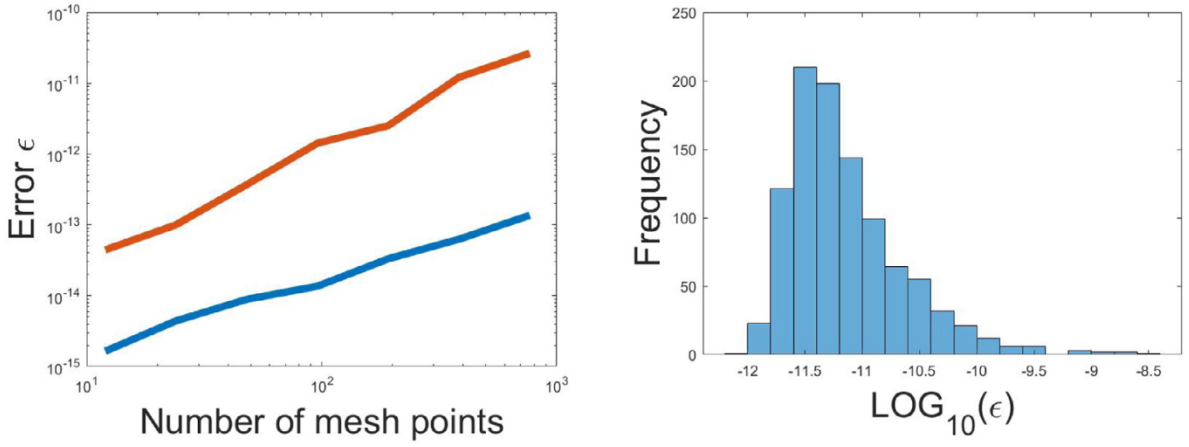

Fig. 2. Left: the behaviour of the error ε as a function of the number of nodal points for $\alpha_1 = \alpha_2 = 1$. The upper red curve is the arithmetic mean of the error of 1000 samples with randomised meshes, the lower blue curve is for equidistant meshes; Right: histogram of the logarithm with basis 10 of the error for randomised meshes after 1000 samples with a mesh of 768 nodes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Numerical examples, discussion and conclusions

5.1. Numerical case studies

In all the simulations, we use $\alpha_1 = \alpha_2 = 1$. For other values of these α–values, the behaviour of the solution is similar. We start with a 2×2 matrix. The mesh points are located on $x = 0$ and on $x = 1$. The resulting finite element discretisation matrix, and its inverse computed from Theorem 3.1 are given by

$$S = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \quad \Longleftrightarrow \quad S^{-1} = \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix},$$

which is consistent with the classical formula for the inversion of a 2×2 matrix, where the entries on the main diagonal are interchanged and the sign of the off diagonals is flipped after division by the determinant. For a discretisation with three unknowns, we obtain the following 3×3 matrix as well as its inverse

$$S = \begin{pmatrix} 3 & -2 & 0 \\ -2 & 4 & -2 \\ 0 & -2 & 3 \end{pmatrix} \quad \Longleftrightarrow \quad S^{-1} = \begin{pmatrix} 2/3 & 1/2 & 1/3 \\ 1/2 & 3/4 & 1/2 \\ 1/3 & 1/2 & 2/3 \end{pmatrix}.$$

We have done this for larger number of nodal points, both in an equidistant mesh and a randomised mesh. We compute the residual matrix

$$R = I - S^{-1}S.$$

This matrix should be zero under infinite precision. Rounding errors will cause deviations from zero. We define the error by

$$\varepsilon = \left(\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} R_{ij}^2 \right)^{1/2}.$$

We have done some simulations with different mesh sizes for both equidistant and randomised meshes. The results can be seen in Fig. 2. It can be seen that for both randomised and equidistant meshes, the error increases with the same order as the size of the matrix increases. This is caused by the propagation of rounding errors, from which we conclude that for larger resolutions, the impact of rounding errors increases. The increase is sharper for randomised meshes. It is believed that this error increase is caused by having a smaller minimal finite element mesh size in a randomised mesh than the mesh size for equidistant meshes using the same number of mesh points. Similar observations were obtained in [19] where the propagation of rounding errors in adaptive finite element meshes was studied. The histogram of the error has been plotted as well in Fig. 2. From the histogram it can be seen that there is some variation in the errors, however, for all meshes that we used, the errors stay within the order of 10^{-10}, which is very small to possible numerical (finite element) errors, which, in case of infinite precision decrease to zero quadratically upon using Lagrangian finite element [illegible]lly confirm the validity of Theorem 3.1.

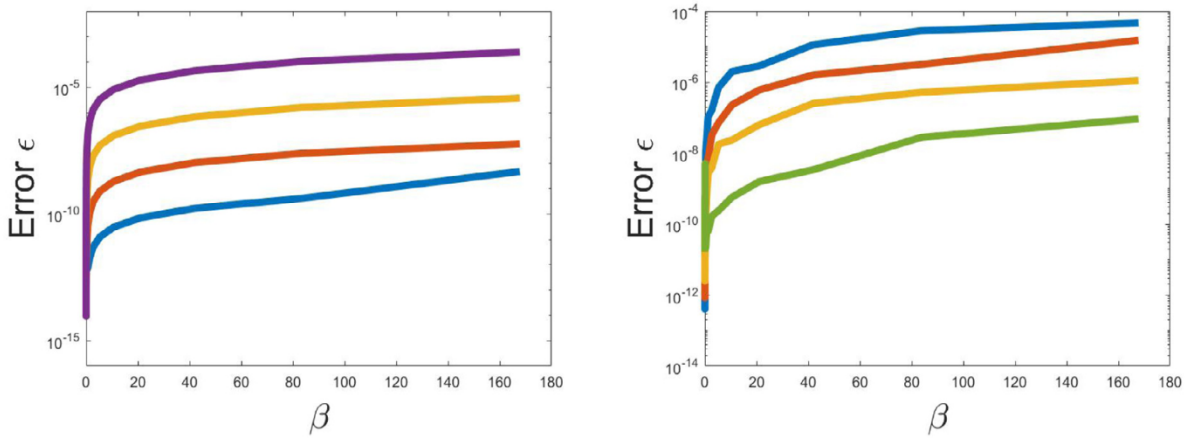

Fig. 3. The behaviour of the error ε as a function of β for $\alpha_1 = \alpha_2 = 1$. Each curve represents a number of finite element nodes: 48, 192, 768 and 3072 nodes (from top to bottom). Left: equidistant mesh; Right: one sample with a randomised mesh.

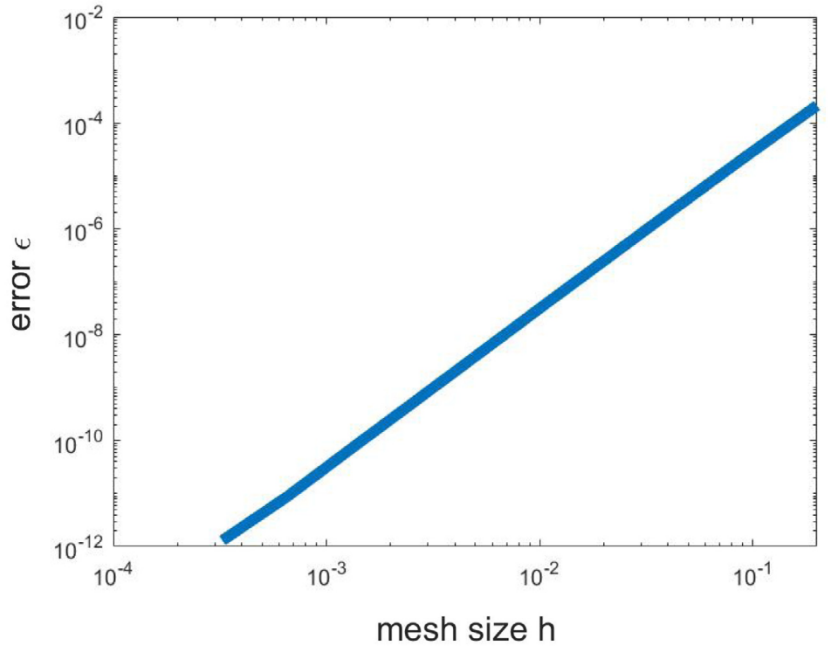

Fig. 4. The error for the approximate inverse as a function of the finite element mesh size for an equidistant mesh for $\alpha_1 = \alpha_2 = 1$ and $\beta = 10$.

Next, we quantitatively analyse the approximate inverse in Theorem 4.1, see also Eq. (18), from the linear steady-state 1D reaction–diffusion Eq. (12). As in the previous problem, we use the values $\alpha_1 = \alpha_2 = 1$ and we take different values of the reaction rate parameter β to study the error of the approximate inverse. We have done simulations for equidistant and randomised meshes, and the results for different values of β have been plotted in Fig. 3. The results are shown for different numbers of mesh points. It is clear that the error increases with increasing value of β. The increase is more or less linear. Furthermore, it can be observed that the error decreases when the number of mesh points increases (hence the mesh size decreases), see also Fig. 4. From Fig. 4, it can be seen that the experimental order of convergence is about 2. This is consistent with Theorem 4.1.

5.2. Discussion

The fundamental solutions to this class of boundary value problems in $\mathbb{R}^1$ can be formulated by

where L denotes a linear operator pending homogeneous boundary conditions. Here G is the fundamental solution. Using linearity of the operator, and the nature of the Dirac delta distribution, one can formally write the solution to a generic boundary value with the same homogeneous boundary conditions

$$Lu = f(x),$$

by convolution of the function $f(x)$ and fundamental solution $G(x;s)$, that is,

$$u(x) = \int_{\Omega} f(s)\, G(x;s) ds,$$

If one uses a finite element method with n unknowns (degrees of freedom) to approximate the solution to the boundary value problem, then one arrives at

$$S\mathbf{c} = \mathbf{b}.$$

The idea in the current study is to express the fundamental solutions that are obtained by letting the Dirac delta distributions act on the finite element mesh points. Subsequently, these fundamental solutions define the inverse of the discretisation matrix. This means that

$$(S^{-1})_{ij} = G(x_i; x_j).$$

Then the i–th entry of the discrete solution can be expressed by

$$c_i = \sum_{j=1}^{n} (S^{-1})_{ij} b_j = \sum_{j=1}^{n} G(x_i; x_j) b_j.$$

This equation gives an explicit formula for the finite element solution with piecewise linear basis functions, and this equation can be applied in general to construct explicit (closed form) expressions for the finite element solution under general functions in the right-hand side and general boundary conditions. As an example, we consider the following boundary value problem:

$$\begin{cases} -u'' = 1, & x \in (0,1), \\ u(0) = 0, & u(1) = 1, \end{cases}$$

then the exact solution is given by

$$u(x) = \frac{1}{2}\, x\,(1-x) + x.$$

Using an equidistant mesh (mesh size h) with n unknowns with piecewise linear basis functions, this gives for the right-hand side

$$b_j = h = \frac{1}{n+1}, \quad \text{for } j = 1, \ldots, n \qquad \text{and } b_n = h + \frac{1}{h} = \frac{1}{n+1} + n + 1.$$

Then using the expression $G(x_i; x_j) = \frac{n+1-j}{n+1} i - (i-j)_+$ at mesh point $x_i = ih$, one recovers the finite element solution by

$$c_i = \sum_{j=1}^{n} G(x_i; x_j) b_j = \sum_{j=1}^{n} \left(\frac{n+1-j}{n+1} i - (i-j)_+ \right) b_j.$$

From this expression, it can be seen that in order to obtain the entire solution vector, that is c_i for $i = 1, \ldots, n$, the amount of work is of the order $\mathcal{O}(n^2)$ floating point operations (flops) (note that S^{-1} is a full matrix). In order to obtain the solution using Gaussian elimination or the LU decomposition, the number of floating point operations is of order $\mathcal{O}(n^3)$, see [20]. However, for a band matrix S, like in this case, the number of flops becomes $\mathcal{O}(n)$ (in particular $\mathcal{O}(2nb)$ where b is half the bandwidth of the matrix). From this point of view, the use of the inverse is not advantageous. The current explicit version of the inverse, or its approximation, allows to compute (approximate) the solution just at the point of interest. That is, the total solution vector does not have to computed. This can be advantageous from a theoretical point of view. We intend to use the insights that were developed in this paper for the analysis of poro-elastic problems with general boundary conditions.

The method of fundamental solutions is feasible as long as the fundamental solutions can be evaluated easily and as long as the resulting finite element error is zero or converges to zero as the mesh size $h \longrightarrow 0$. For higher dimensional cases, the fundamental solution may be expressed in terms of Fourier series for rectangular geometries in $\mathbb{R}^2$, which reads as

$$4 \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{\sin(n\pi x)\sin(m\pi y)\sin(n\pi x_j)\sin(m\pi y_j)}{n^2+m^2}.$$

The above equation represents the solution to the Poisson equation on a unit rectangle with a pulse on $\mathbf{x}_j = (x_j, y_j)$. An alternative representation is written in terms of the sum of the Green's function, being

$$G^{\infty}(\mathbf{x}; \mathbf{x}_j) = -\frac{1}{2\pi} \ln(\|\mathbf{x} - \mathbf{x}_j\|),$$

for the infinite case, and the solution to the Laplace equation with the opposite sign of the Green's function for the infinite case on the boundary, which is defined by

$$\begin{cases} -\Delta v = 0, \text{ in } (0, 1)^2, \\ v|_{\Gamma} = -G^{\infty}(\mathbf{x}; \mathbf{x}_j). \end{cases}$$

Here Γ represents the boundary of $(0, 1)^2$. Then $G(\mathbf{x}; \mathbf{x}_j) = G^{\infty}(\mathbf{x}; \mathbf{x}_j) + v$, which is just a different representation of G. The latter approach may be useful in determining finite element solutions. In either case, these solutions involve infinite sums, and therewith they are somewhat more elaborate. Furthermore, in practice, these solution formulas do not allow an accessible representation (or approximation) of the inverse. For this reason, this has not been worked out in the current study.

For the sake of solving systems of linear equations, inversion of the system (discretisation) matrix is extremely expensive. The current one-dimensional case gives a tridiagonal matrix. Tridiagonal matrices can be treated very efficiently by the use of the Thomas algorithm. Since the inversion of matrices is very expensive and needs the solution of n systems of linear equations, one hardly ever determines the inverse of a matrix. A possible exception could be where one needs to solve a huge number of systems of linear equations where the same system matrix is used in every solve. On the other hand, the inverse of a matrix could be important for theoretical reasons where one aims at proving certain properties of discretisations of complicated systems. Important examples are Biot's or Terzaghi's system for poroelasticity, where finite element solutions may loose monotonicity properties and where stabilisation may be needed to overcome spurious oscillations. The inverse of the Laplace matrix may help derive criteria for the mesh to have monotonic solutions.

Regarding elliptic equations with the Dirac delta distribution as the right hand side, finite element solutions have been analysed in several studies in $\mathbb{R}^d$. Scott [21] proved that the L^2 norm of the finite element error satisfies $\|u - u_h\|_{L^2(\Omega)} \leq Ch^{2-d/2}$ for linear Lagrangian elements. Later Erikson [22] extended the result to $\|u - u_h\|_{L^1(\Omega)} \leq Ch^{k+1}$ for $\mathbb{P}_k$ elements in two spatial dimensions ($d = 2$). Apel et al. [23] proved that $\|u - u_h\|_{L^2(\Omega)} \leq Ch^2\sqrt{|\log(h)|}$ for linear Lagrangian elements in $\mathbb{R}^2$. D'Angelo [24] demonstrated stability and convergence of finite element solutions in weighted Sobolev spaces to Dirac delta problems. Köppl and Wohlmuth [25] proved convergence of the L^2-error away from the singularity, that is, away from the location of action of the Dirac delta distribution. Their main result can be summarised by $\|u - u_h\|_{L^2(\Omega_0)} \leq Ch^{k+1}$ where $\Omega_0 \subset \Omega$ is a proper open subset of Ω that does not contain the singularity for $\mathbb{P}_k$–elements. Their proof was extended to the finite element error in the H^1 norm away from the singularity by Bertulozza et al. [26]. All this work was done in $\mathbb{R}^2$. The result by Wahlbin [18] does not express the error in a Sobolev norm, however, it does in a max norm, and it is only valid under strict regularity conditions, which do not hold in our case. Furthermore, the result by [18] has been obtained for two dimensions. Since the results of the current paper estimate the error of the inverse Laplacian for a linear one-dimensional reaction–diffusion equation with generic boundary conditions, it also estimates the point wise error of the finite element solution to $\mathcal{O}(h^{3/2})$. From this it can be concluded that the finite element error of the max norm over the mesh points is also given by $\mathcal{O}(h^{3/2})$.

5.3. Conclusions

We have illustrated that the use of fundamental solutions is a straightforward procedure to derive the inverse of the one-dimensional Laplace matrix. This technique can be applied for general homogeneous boundary conditions, hence for combinations of Dirichlet, Neumann and Robin boundary conditions. Furthermore, this procedure can be applied to derive an approximate inverse of a linear 1D reaction-Laplace equation, which is an approximation to the real inverse. It has been demonstrated that the approximate inverse converges to the inverse as the mesh size, h, tends to zero.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Davis TA. Direct methods for sparse linear systems. Society of Industrial and Applied Mathematics; 2006.
[2] Strang G. Linear algebra and its applications. 2nd ed.. Academic Press; 1980.
[3] Quarteroni A, Sacco R, Saleri F. Numerica mathematics. Texts in applied mathematics, vol. 37, Springer; 2000.
[4] Golub GH, van Loan CF. Matrix computations. 3rd ed.. The John Hopkins University Press; 1996.
ltigrid. Academic Press; 2000.
luid dynamics. Springer series in computational mathematics, vol. 29, Springer; 2000.

[7] Axelsson O. Iterative solution methods. Cambridge University Press; 1996.
[8] van der Vorst HA. Iterative Krylov methods for large linear systems. Cambridge University Press; 2003.
[9] Nabben R, Vuik C. A comparison of deflation and coarse grid correction applied to porous media flow. SIAM J Numer Anal 2004;42(4):1631–47.
[10] Gueye SB. The exact formulation of the inverse of the tridiagonal matrix for solving the 1D Poisson equation with the finite difference method. J Electromagn Anal Appl 2014;6(10):303–8.
[11] Evans LC. Partial differential equations. Graduate studies in mathematics, vol. 19, American Mathematical Society; 2002.
[12] Borregales M, Kumar K, Nordbotten JM, Radu FA. Iterative solvers for Biot model under small and large deformation. Comput Geosci 2021;25:687–99.
[13] Hu X, Mu L, Ye X. Weak Galerkin method for the Biot's consolidation model. Comput Math Appl 2018;75:2017–30.
[14] Rodrigo C, Gaspar FJ, Hu X, Zikatanov LT. Stability and monotonicity for some discretisations of the Biot's consolidation model. Comput Methods Appl Mech Engrg 2016;298:193–204.
[15] Brenner SC, Scott LR. The mathematical theory of finite element methods. Texts in applied mathematics, 3rd ed.. vol. 15, Springer; 2008.
[16] Atkinson K, Han W. Theoretical numerical analysis: a functional analysis framework. Texts in applied mathematics, 3rd ed.. vol. 39, Springer; 2009.
[17] Kashiwabara T, Kemmochi T. Maximum norm error estimates for the finite element approximation of parabolic problems on smooth domains. 2018, arXiv:1805.01336.
[18] Wahlbin LB. Maximum norm error estimates in the finite element method with isoparametric quadratic elements and numerical integration. RAIRO Anal Numér 1978;12(2):173–202.
[19] Alvarez-Aramberri J, Pardo D, Paszynski M, Collier N, Dalcin L, Calo V. On round–off error for adaptive finite element methods. Procedia Comput Sci 2012;9:1474–83.
[20] Vuik C. In: Burgerscentrum JM, editor. Iterative solution methods. Lecture notes at research school for fluid mechanics, The Netherlands: Delft University of Technology; 2021.
[21] Scott LR. Finite element convergence for singular data. Numer Math 1978;21:317–27.
[22] Erikson K. Finite element methods of optimal order for problems with singular data. Math Comp 1985;44(170):345–60.
[23] Apel T, Benedix D, Sirch D, Vexler B. A priori mesh grading for an elliptic problem with Diract delta source terms. SIAM J Numer Anal 2011;49(3):992–1005.
[24] D'Angelo C. Finite element approximation of elliptic problems with Dirac measure terms in weighted spaces: applications to one-and three-dimensional coupled problems. SIAM J Numer Anal 2012;50(1):194–215.
[25] Köppl T, Wohlmuth B. Optimal a priori error estimates for an elliptic problem with Dirac right–hand side. SIAM J Numer Anal 2014;52(4):1753–69.
[26] Bertoluzza S, Decoene A, Lacouture L, Martin S. Local error estimates of the finite element method for an elliptic problem with a Dirac source term. 2015, Hal-01150745v5 preprint.