

## FAILURE OF THE DISCRETE MAXIMUM PRINCIPLE FOR AN ELLIPTIC FINITE ELEMENT PROBLEM

ANDREI DRĂGĂNESCU, TODD F. DUPONT, AND L. RIDGWAY SCOTT

ABSTRACT. There has been a long-standing question of whether certain mesh restrictions are required for a maximum condition to hold for the discrete equations arising from a finite element approximation of an elliptic problem. This is related to knowing whether the discrete Green's function is positive for triangular meshes allowing sufficiently good approximation of  $H^1$  functions. We study this question for the Poisson problem in two dimensions discretized via the Galerkin method with continuous piecewise linears. We give examples which show that in general the answer is negative, and furthermore we extend the number of cases where it is known to be positive. Our techniques utilize some new results about discrete Green's functions that are of independent interest.

### 1. INTRODUCTION

Let  $D \subset \mathbb{R}^2$  be a bounded polygonal domain. We consider the *Poisson problem*

$$(1.1) \quad \begin{cases} -\Delta u = f, & \text{on } D, \\ u = \tilde{u}, & \text{on } \partial D, \end{cases}$$

where  $\tilde{u} \in C^0(\partial D)$  and  $f$  is in an appropriate space to be described later. It is an elementary fact that the following *maximum principle* – which we shall refer to occasionally as the *boundary maximum principle* – holds for the homogeneous problem ( $f = 0$ ):

$$(1.2) \quad \|u\|_{\infty, D} \leq \|\tilde{u}\|_{\infty, \partial D}.$$

This can be regarded as a consequence of the positivity of the *Green's function*  $G(x, y)$ :

$$(1.3) \quad G_y(x) > 0, \quad \text{for all } x \in D,$$

where  $G_y(\cdot) = G(\cdot, y)$  solves (1.1) with  $f = \delta_y$  (point mass at  $y$ ) and  $G_y|_{\partial D} = 0$ . See [6] for results regarding the positivity of the Green's functions for elliptic problems.

---

Received by the editor May 12, 2003 and, in revised form, June 8, 2003.

2000 *Mathematics Subject Classification*. Primary 65N30; Secondary 65N50.

*Key words and phrases*. Finite elements, Green's function, discrete maximum principle.

The work of the authors was supported by the ASCI Flash Center at the University of Chicago under DOE contract B532820 and by the MRSEC Program of the National Science Foundation under award DMR-0213745.

Given a triangulation  $\mathcal{T}_h$  of  $D$ , we define the following spaces:

$$(1.4) \quad \mathcal{M}_h = \mathcal{M}(\mathcal{T}_h) = \{u \in \mathcal{C}^0(\bar{D}) : u|_T \text{ is linear for each } T \in \mathcal{T}_h\},$$

$$(1.5) \quad \mathcal{M}_{0,h} = \mathcal{M}_0(\mathcal{T}_h) = \{u \in \mathcal{M}_h : u|_{\partial D} = 0\}.$$

The discrete problem associated to (1.1) via the Galerkin method is: Find  $u_h \in \mathcal{M}_h$  such that

$$(1.6) \quad \begin{cases} a(u_h, v) = \langle f, v \rangle, & \text{for all } v \in \mathcal{M}_{0,h}, \\ u_h - \tilde{u}_h \in \mathcal{M}_{0,h}, \end{cases}$$

where

$$(1.7) \quad a(u, v) = \int_D \nabla u \cdot \nabla v,$$

$\tilde{u}_h$  is a given function in  $\mathcal{M}_h$  and  $\langle \cdot, \cdot \rangle : \mathcal{M}_h^* \times \mathcal{M}_h \rightarrow \mathbb{R}$  is the dual pairing. By analogy with the continuous case we denote by  $G_y^h$  the *discrete Green's function*, namely the solution of (1.6) with  $\tilde{u}_h = 0$  and  $f = \delta_y$ .

Let  $P_i, 1 \leq i \leq n$  (resp.  $n+1 \leq i \leq n+m$ ), be the vertices of the triangulation which belong to the interior (resp. boundary) of  $D$ , and let  $\mathcal{B} = \{\phi_i \in \mathcal{M}_h : 1 \leq i \leq n+m\}$  be the *nodal basis* for  $\mathcal{M}_h$  (i.e.,  $\phi_i(P_j) = \delta_{ij}$ ). With this notation, if  $u \in \mathcal{M}_{0,h}$  then  $u = \sum_{i=1}^n u_i \phi_i$  and  $u(P_i) = u_i$ ; hence  $u$  can be represented as the vector  $\mathbf{U} = (u_1, u_2, \dots, u_n)^t$ . The system (1.6) is rewritten in matrix form as

$$(1.8) \quad \mathbf{A} \mathbf{U} = \mathbf{B}$$

where  $\mathbf{A} = (a_{ij})_{1 \leq i, j \leq n}$  is the stiffness matrix, i.e.,  $a_{ij} = a(\phi_j, \phi_i)$ , and the right-hand side  $\mathbf{B} = (b_1, \dots, b_n)^t$  is given by

$$(1.9) \quad b_i = \langle f, \phi_i \rangle - \sum_{j=n+1}^{n+m} \tilde{u}_h(P_j) a_{ij}.$$

As in the continuous case, we address mainly the question of whether

$$(1.10) \quad G_y^h(x) > (\text{or } \geq) 0, \text{ for all vertices } x, y \text{ in the triangulation.}$$

We discuss the implication “(1.10)  $\Rightarrow$   $u_h$  satisfies the boundary maximum principle (1.2)” in Section 4.

**Definition 1.1.** We call an edge  $\mathbf{e} \notin \partial D$  of the triangulation  $\mathcal{T}_h$  **nonnegative** if the sum of the two angles opposite to it is  $\leq \pi$ ; otherwise it is **negative**.

The discrete system for  $-\Delta$  can be viewed as a resistance network, where each edge is replaced by a resistor. It will turn out (cf. Corollary A.2) that the conductivity (the reciprocal of the resistance) of a given edge  $\mathbf{e}$  is nonnegative iff  $\mathbf{e}$  is a nonnegative edge.

Matrix theoretic considerations give rise to the following well-known condition which implies (1.10):

**Condition 1.2** (Interior edge condition). All interior edges are nonnegative. (By “interior” we mean that **neither** of its vertices lies on the boundary.)

Condition 1.2 is implied by an earlier version (“*all angles are  $\leq \pi/2$* ”) which appears in [2], and it is found as formulated on page 78 in [10] and other articles on the subject that followed (to select just a few: [4], [7], [12], [3]). A more relaxed version of Condition 1.2 has been obtained in [7]; it allows for the existence of slightly negative edges (the sum of the opposite angles has to be bounded by  $\pi + \epsilon$

for some  $\epsilon > 0$ ) but adds restrictions involving a larger neighborhood of each edge. This result has been extended to three dimensions in [5].

Other versions of a discrete maximum principle are discussed in the literature: first one can ask whether – as in the continuous case – discrete harmonic functions satisfy

$$(1.11) \quad \|u\|_{\infty, D'} \leq \|\tilde{u}\|_{\infty, \partial D'}, \quad \text{for every mesh – subdomain } D'.$$

It is shown in [4] that for higher order Lagrange elements the maximum principle (1.11) holds only under very strong geometrical constraints. Finally, in [8] a weaker discrete maximum principle is shown to hold under quite general conditions on the mesh (quasi-uniformity) and arbitrary degree polynomials, namely

$$(1.12) \quad \|u\|_{\infty, D} \leq C \|\tilde{u}\|_{\infty, \partial D},$$

where  $C > 0$  is independent of the resolution  $h$ .

One critical question we address here is whether conditions on the mesh other than quasi-uniformity are needed in order for (1.10) to hold. The discussion at the end of Section 2 and the results in Section 3 suggest that (1.10) might be true if only edges near the boundary are nonnegative. In Section 4 we show that in order for the boundary maximum principle to hold for discrete harmonic functions, in addition to the Green's function being nonnegative the nonnegativity of the nonboundary edges with one vertex on  $\partial D$  is relevant. It is not hard to find isolated situations where (1.10) fails even on a large set; e.g., in Example 2.2(ii) the Green's function is negative at half of the interior nodes, and there are known cases (see [7] and Example 3.3(i)) for which it has been shown computationally that the Green's function has nonnegative values as the mesh-size goes to zero even if the mesh is *quasi-uniform*. In Section 5 we give an example where *only one* edge near  $\partial D$  violates Condition 1.2 and (1.10) fails, again, on a quasi-uniform mesh. Our techniques involve some new results regarding discrete Green's functions that are of independent interest, e.g., Theorem 5.7. Finally, in Section 6 we show that (1.10) holds for a slightly more general class of meshes.

## 2. PRELIMINARIES

For the sake of completeness we provide the arguments leading to the implication

$$(2.1) \quad \text{Condition 1.2} \Rightarrow (1.10).$$

For  $y = P_j$  the vector  $\mathbf{G}^{(j)} = (G_y^h(P_1), \dots, G_y^h(P_n))^t$  solves

$$(2.2) \quad \mathbf{A} \mathbf{G}^{(j)} = \mathbf{B}^{(j)},$$

where

$$(2.3) \quad \mathbf{B}_i^{(j)} = \langle \delta_{P_j}, \phi_i \rangle = \phi_i(P_j) = \delta_{ij},$$

i.e.,  $(\mathbf{B}^{(j)})_{1 \leq j \leq n}$  is the standard basis for  $\mathbb{R}^n$ . If we regard the vectors as column matrices, then the previous equations can be compactly written

$$(2.4) \quad \mathbf{A} \left[ \mathbf{G}^{(1)}, \dots, \mathbf{G}^{(n)} \right] = \left[ \mathbf{B}^{(1)}, \dots, \mathbf{B}^{(n)} \right] = \mathbf{I}_n,$$

which shows that

$$(2.5) \quad G_{P_j}(P_i) = (\mathbf{A}^{-1})_{ij}.$$

In other words, *the values of the discrete Green's function are the corresponding entries in the inverse of the stiffness matrix*  $\mathbf{A} = (a_{ij})$  if we use the nodal basis.

Throughout this paper we use the following convention: for a matrix  $\mathbf{B} = (b_{ij})$ , the notation  $\mathbf{B} >$  (*resp.*  $\geq, <, \leq$ )  $\mathbf{0}$  means  $b_{ij} >$  (*resp.*  $\geq, <, \leq$ )  $0$  for all  $i, j$ , and this applies to vectors also; furthermore, we write  $\mathbf{B} \gneq \mathbf{0}$  for ( $\mathbf{B} \geq \mathbf{0}$  and  $\mathbf{B} \neq \mathbf{0}$ ).

*Remark 2.1.* If  $\mathbf{B}$  is a matrix and  $\mathbf{r}$  is a vector, then

- (a)  $\mathbf{B} > \mathbf{0}$  iff  $\mathbf{B} \mathbf{s} > \mathbf{0}$  for all vectors  $\mathbf{s} \gneq \mathbf{0}$ ;
- (b) if  $\mathbf{B} \gneq \mathbf{0}$  and  $\mathbf{r} > \mathbf{0}$ , then  $\mathbf{B} \mathbf{r} \gneq \mathbf{0}$ .

A sufficient condition on a symmetric matrix  $\mathbf{A} = (a_{ij})$  for  $\mathbf{A}^{-1} \gneq \mathbf{0}$  is that  $\mathbf{A}$  be a *Stieltjes-matrix*; that is, if  $\mathbf{A}$  is **positive definite** and

$$(2.6) \quad a_{ij} \leq 0 \text{ for } i \neq j,$$

then  $\mathbf{A}^{-1} \gneq \mathbf{0}$  (cf. Corollary 3.24 in [11]). If in addition  $\mathbf{A}$  is irreducible, then  $\mathbf{A}^{-1} > \mathbf{0}$ . Since the stiffness matrix  $\mathbf{A}$  is positive definite and Corollary A.2 implies the equivalence of Condition 1.2 and (2.6), the implication (2.1) now holds.

The following argument suggests that (1.10) holds for more general meshes, and it is quite likely that the sort of angle condition described here is **not** essential. Local error estimates (cf. [9]) indicate that  $|G_y(x) - G_y^h(x)|$  should be small for  $|x - y|$  not too small. Similarly, if  $G_y(x)$  is really large (say for  $y$  in the interior), then the same local techniques imply that  $G_y^h(x)$  cannot be near zero. Therefore  $G_y^h(x)$  should be positive for

- (1)  $y$  in the interior and  $x$  arbitrary.

For  $y$  near the boundary (where one assumes  $G_y$  vanishes, i.e. Dirichlet conditions), the fact that the gradient of  $G_y$  does not vanish on the boundary (see (4.2)) implies that  $G_y^h(x)$  must not vanish if  $|x - y|$  is not small, again using local error estimates as in [9] (one can show that the gradients of  $G_y$  and  $G_y^h$  are sufficiently close). Thus  $G_y^h(x)$  should be positive for

- (2)  $y$  near the boundary and  $|x - y|$  not small.

This leaves the case

- (3)  $y$  near the boundary and  $|x - y|$  small.

In case (3) the above simple arguments appear to break down. This suggests that a good place to look for a counterexample for (1.10) should be in case (3) above and where at least one edge near the boundary is negative.

In Section 3 we show that, under assumptions which follow naturally from the discussion above, only edges near the boundary need to be nonnegative for (1.10) to hold. Section 5 contains examples of a nonpositive discrete Green's function and a nonpositive discrete harmonic function with nonnegative boundary values, essentially showing that the result in Section 3 is sharp.

We would like to add some remarks on the irreducibility of the stiffness matrix  $\mathbf{A}$ . One thing to note is that  $\mathbf{A}$  may not be irreducible even if the mesh is connected, as shown in the next example.

**Example 2.2.** Let the domain consist of two unit size squares with sides parallel to the coordinate axes and joined together by a narrow corridor. More precisely let  $D$  be the interior of the set:

$$D = ([-1 - h_1, -h_1] \cup [h_1, 1 + h_1]) \times [-1, 1] \cup [-h_1, h_1] \times [-h, h],$$

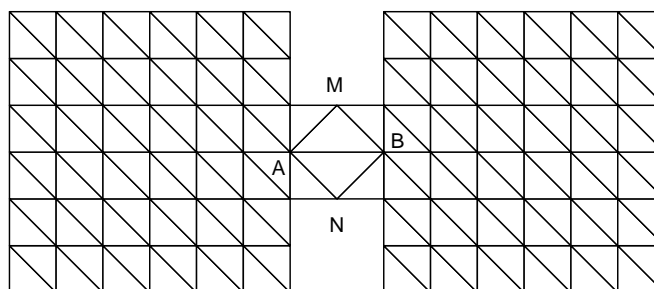


FIGURE 1. Here  $h_1 = h$  and the stiffness matrix is reducible for this mesh as  $a(\phi_A, \phi_B) = 0$ .

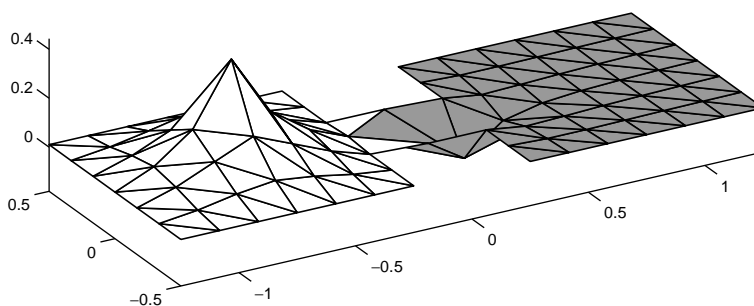


FIGURE 2. All values in the interior of the right square of the Green’s function with source in the center of the left square are negative; here  $h_1 = 3h/2, h = 1/6$ . Triangles where at least one value is negative are grey; negative values are magnified 100 times for enhanced visibility.

where  $(2h)^{-1} \in \mathbb{N}$  and  $h_1 > 0$ , and consider on each of the two unit size squares the uniform triangular mesh of size  $h$ , while the corridor is triangulated as in Figure 1.

- (i) If  $h_1 = h$ , then  $\angle AMB = \angle ANB = \pi/2$  (refer to Figure 1), and by (A.2) we get  $a(\phi_A, \phi_B) = 0$ ; therefore the stiffness matrix has a block structure,  $\mathbf{A} = \text{blockdiagonal}(\mathbf{A}_l, \mathbf{A}_r)$ , where  $\mathbf{A}_l$  (resp.  $\mathbf{A}_r$ ) is the stiffness matrix corresponding to the union between the left (resp. right) square and  $[-h, h] \times [-h, h]$ .
- (ii) If  $h_1 > h$ , then  $a(\phi_A, \phi_B) > 0$ ; hence the edge  $AB$  is negative. Numerical experiments show that by placing a unit mass forcing term at some node in the left square, all values of the solution to (1.6) with zero boundary conditions are negative in the right square (see Figure 2).

It should be noted that the domain in Example 2.2(i) depends on  $h$ ; if we keep  $h$  fixed and refine the mesh for example by cutting each triangle in four equal pieces similar to the original, then the reducibility goes away.

The difficulty of constructing a “normal” mesh that gives a reducible stiffness matrix may be related to the following remark.

*Remark 2.3.* If the mesh is connected, then the stiffness matrix that includes the boundary nodes

$$\tilde{\mathbf{A}} = (a(\phi_j, \phi_i))_{1 \leq i, j \leq m+n},$$

which is related to the Neumann problem, is irreducible.

*Proof.* Suppose  $\tilde{\mathbf{A}}$  is reducible. Since it is symmetric, this implies it can be written as

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix}.$$

Let  $u \in \mathcal{M}_h$  be defined to be 1 at the nodes corresponding to the rows in  $\mathbf{A}_1$  and 0 at all the other nodes, and let  $\mathbf{U} = (u(P_1), \dots, u(P_{m+n}))^t$ . Since

$$\sum_{j=1}^{m+n} a(\phi_j, \phi_i) = 0, \text{ for all } 1 \leq i \leq m+n,$$

we get  $\mathbf{A}\mathbf{U} = \mathbf{0}$ ; hence

$$\int \nabla u \cdot \nabla u = \mathbf{U}^t \mathbf{A} \mathbf{U} = 0.$$

This makes  $u$  constant on each triangle, and the connectivity of the mesh now implies that  $u$  is a constant, thus contradicting the assumption.  $\square$

### 3. REDUCTION OF THE INTERIOR EDGE CONDITION TO CLOSE-TO-BOUNDARY EDGES

The main result of this section—Theorem 3.2—shows that if we are able to prove by other means (e.g., along the lines of the heuristic arguments in the previous section) that for a quite general class of meshes the discrete Green's function is positive for singularities  $x \in D$  lying *away* from  $\partial D$ , and if we assume that the edges near the boundary are nonnegative, then the discrete Green's function is positive everywhere.

We start with a preliminary

**Lemma 3.1.** *Let  $\mathbf{A} \in \mathbb{R}^{n \times n}$  be a symmetric nonsingular matrix having the following structure (which corresponds to a splitting  $\mathbb{R}^n = \mathbb{R}^{n_1} \oplus \mathbb{R}^{n_2}$ ):*

$$(3.1) \quad \mathbf{A} = \begin{bmatrix} \mathbf{e} & \mathbf{H} \\ \mathbf{H}^t & \mathbf{F} \end{bmatrix},$$

with  $\mathbf{H} \leq \mathbf{0}$  (resp.  $\mathbf{H} \not\leq \mathbf{0}$ ),  $\mathbf{e}$  nonsingular and  $\mathbf{e}^{-1} \geq$  (resp.  $>$ )  $\mathbf{0}$ . We further assume that for all  $\mathbf{f} \in \mathbb{R}^{n_2}$

$$(3.2) \quad \mathbf{f} \geq \mathbf{0} \text{ (resp. } \mathbf{f} \not\geq \mathbf{0}) \text{ implies } \mathbf{A}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{f} \end{bmatrix} \geq \text{ (resp. } >) \mathbf{0}.$$

Then  $\mathbf{A}^{-1} \geq$  (resp.  $>$ )  $\mathbf{0}$ .

*Proof.* We prove only the strict inequalities, i.e., the ones in parentheses, since the proof in the nonstrict version goes along the same lines. Let  $\mathbf{b}^t = (\mathbf{e}^t, \mathbf{f}^t) \in \mathbb{R}^{n_1} \oplus \mathbb{R}^{n_2}$  be a nonnegative vector, and let  $\mathbf{x}^t = (\mathbf{u}^t, \mathbf{v}^t) \in \mathbb{R}^{n_1} \oplus \mathbb{R}^{n_2}$  be the solution to

$$(3.3) \quad \mathbf{A} \mathbf{x} = \mathbf{b}.$$

We rewrite the system above as

$$(3.4) \quad \begin{cases} \mathbf{E} \mathbf{u} + \mathbf{H} \mathbf{v} &= \mathbf{e}, \\ \mathbf{H}^t \mathbf{u} + \mathbf{F} \mathbf{v} &= \mathbf{f}, \end{cases}$$

which is equivalent to

$$(3.5) \quad \mathbf{u} = \mathbf{e}^{-1}(\mathbf{e} - \mathbf{H} \mathbf{v}),$$

$$(3.6) \quad (\mathbf{F} - \mathbf{H}^t \mathbf{e}^{-1} \mathbf{H}) \mathbf{v} = \mathbf{f} - \mathbf{H}^t \mathbf{e}^{-1} \mathbf{e}.$$

Condition (3.2) implies that if  $\mathbf{f} \succeq \mathbf{0}$  and  $\mathbf{e} = \mathbf{0}$ , then  $\mathbf{x} > \mathbf{0}$ , in particular  $\mathbf{v} > \mathbf{0}$ . Since this is true for any  $\mathbf{f} \succeq \mathbf{0}$ , Remark 2.1(a) now implies that

$$(3.7) \quad (\mathbf{F} - \mathbf{H}^t \mathbf{e}^{-1} \mathbf{H})^{-1} > \mathbf{0}.$$

If we now take  $\mathbf{f} = \mathbf{0}$  and  $\mathbf{e} \succeq \mathbf{0}$ , the hypotheses and Remark 2.1(a) again imply that  $\mathbf{e}^{-1} \mathbf{e} > \mathbf{0}$ , and by Remark 2.1(b) and  $-\mathbf{H} \succeq \mathbf{0}$  we get that  $-\mathbf{H}^t \mathbf{e}^{-1} \mathbf{e} \succeq \mathbf{0}$ . Now by (3.7) we get  $\mathbf{v} > \mathbf{0}$ . Thus  $-\mathbf{H} \mathbf{v} \succeq \mathbf{0}$ , and finally  $\mathbf{u} > \mathbf{0}$  follows by (3.5) and  $\mathbf{e}^{-1} > \mathbf{0}$ .  $\square$

Suppose the collection of indices corresponding to *interior* nodes of the triangulation is divided into two disjoint sets

$$(3.8) \quad \{1, \dots, n\} = I_{ext} \cup I_{int}.$$

Motivated by the discussion in the previous section, the notation above is intended to suggest the interpretation:  $i \in I_{ext}$  (resp.  $I_{int}$ ) means “ $P_i$  is *close to* (resp. *far from*)  $\partial D$ ”, but we do not make any formal assumption related to this meaning. Let  $\mathbf{A}$  be the “global” stiffness matrix. Following the index splitting (3.8), we get a corresponding splitting of  $\mathbf{A}$  as in Lemma 3.1, with  $\mathbf{e}$  (resp.  $\mathbf{F}$ ) being the stiffness matrix associated to the basis elements  $\phi_i$  with  $i \in I_{ext}$  (resp.  $i \in I_{int}$ ) and  $\mathbf{H}$  the matrix containing the other entries:

$$(3.9) \quad \mathbf{H}_{ij} = a(\phi_j, \phi_i), \quad \text{with } i \in I_{ext}, j \in I_{int}.$$

**Theorem 3.2.** *Suppose all edges with one vertex in  $I_{ext}$  and the other one either in  $I_{ext}$  or in  $I_{int}$  are nonnegative in the sense of Definition 1.1.*

- (i) *If  $G_x^h \geq 0$  for all  $x \in I_{int}$ , then  $G_x^h \geq 0$  for all  $x \in D$ .*
- (ii) *If  $G_x^h > 0$  for all  $x \in I_{int}$  and the matrices  $\mathbf{A}$  and  $\mathbf{e}$  are irreducible, then  $G_x^h > 0$  for all  $x \in D$ .*

*Proof.* Again we only prove (ii) as (i) is similar. The hypothesis and Corollary A.2 imply  $\mathbf{H} \leq \mathbf{0}$  and  $\mathbf{e}$  has nonpositive nondiagonal entries, which makes  $\mathbf{e}$  an irreducible Stieltjes matrix; hence  $\mathbf{e}^{-1} > \mathbf{0}$ . Since  $\mathbf{A}$  is irreducible, it follows that  $\mathbf{H} \neq \mathbf{0}$ . Since “ $G_x^h > 0$  for all  $x \in I_{int}$ ” translates into (3.2) in matrix language (cf. Remark 2.1(a)), Lemma 3.1 now applies, and we conclude that  $\mathbf{A}^{-1} > \mathbf{0}$ .  $\square$

The heuristic argument in Section 2 and the usefulness of Theorem 3.2 are supported by the numerical experiments in Example 3.3(ii). First we revisit an example from [7]:

**Example 3.3.** Let  $\mathbf{a} = e^{i\frac{\theta}{2}}$ ,  $\mathbf{b} = e^{-i\frac{\theta}{2}}$  (so that  $\angle(\mathbf{a}, \mathbf{b}) = \theta$ ) and let  $D$  be the convex hull of the rhombus with corners  $\mathbf{0}$ ,  $\mathbf{b}$ ,  $\mathbf{a} + \mathbf{b}$ ,  $\mathbf{a}$ .

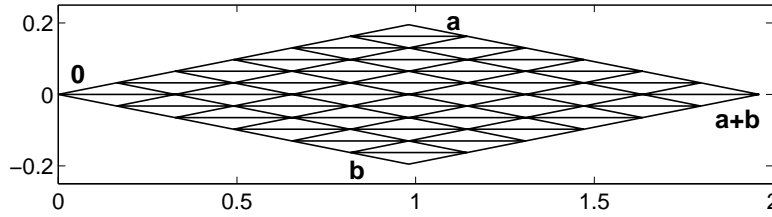


FIGURE 3. The mesh  $\mathcal{T}_6$ ; the sharp angle of the rhombus is  $\theta = \pi/8$ .

- (i) We consider the three line mesh  $\mathcal{T}_n$  on  $D$  obtained by first cutting  $D$  with equidistant lines parallel to  $\mathbf{a}$  and  $\mathbf{b}$ , respectively, and then by slicing each of the  $n^2$  resulting small rhombi along its long diagonal (the one parallel to the  $x$ -axis; see Figure 3). The mesh is obviously quasi-uniform. Condition 1.2 is violated as approximately  $1/3$  of the edges are negative. It is shown in [7] that there exists  $\theta_0 \in (0, \pi/2)$  such that for  $\theta \in (\theta_0, \pi/2]$  the corresponding Green's function  $G^n$  is positive for all  $n > 0$  (however, the discussion in Section 4 implies that the boundary maximum principle is violated if  $\theta < \pi/2$ ). Furthermore it is stated that for  $\theta < \theta_0 = \pi/4$  there are cases where the Green's function is nonpositive. Our computations show that, indeed, for  $\theta = \frac{\pi}{4}$  we have  $\mu(n) = \min_{x,y \in D} G_x^n(y) < 0$  but  $\lim_{n \rightarrow \infty} \mu(n) = 0$ , and for  $\theta < \frac{\pi}{4}$  there are strong indications that  $\lim_{n \rightarrow \infty} \mu(n) < 0$ . Hence this is a natural example of a nonpositive Green's function on a quite “tame” quasi-uniform mesh (see also Example 3.4).
- (ii) We modify the mesh  $\mathcal{T}_n$  from (i) by transforming all negative edges near the boundary into positive ones. More precisely let  $0 \leq k \leq n/2$  be an integer; we obtain the mesh  $\mathcal{T}_{n,k}$  by doing the initial partitioning into  $n^2$  rhombi as in (i) and then cutting all rhombi in the first  $k$  layers near the boundary along their *short* diagonal and the rest of the rhombi along their long diagonal (see Figure 4). Denote by  $G^{n,k}$  the Green's function of the resulting grid and by

$$\mu(n, k) = \min_{x,y \in D} G_x^{n,k}(y).$$

From (i) we know that if  $\theta < \theta_0$ , then  $\mu(n, 0) < 0$ . If  $k = n/2$ , then Condition 1.2 is satisfied; hence  $\mu(n, n/2) = 0$  ( $G^{n,k}$  is zero on  $\partial D$ ). Denote by

$$k(n) = \min\{0 \leq k \leq n/2 : \mu(n, k) = 0\};$$

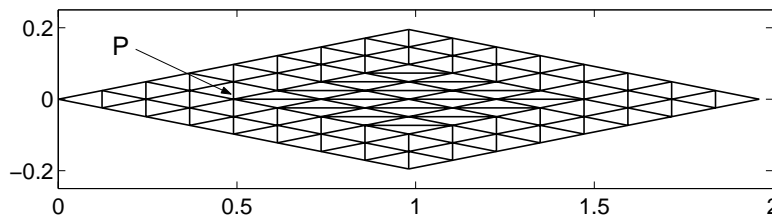


FIGURE 4. The mesh  $\mathcal{T}_{8,2}$ ; the sharp angle of the rhombus is  $\theta = \pi/8$ .

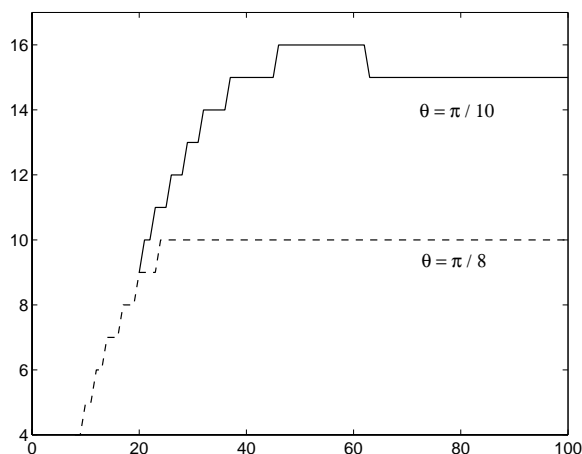


FIGURE 5. The evolution of  $k(n)$  for  $\theta = \pi/10, \pi/8$ .

that is,  $k(n)$  is the width of the thinnest layer for which the Green's function is nonnegative. Our goal is to compute the function  $k(n)$ . Numerical experiments performed with  $\theta = \frac{\pi}{8}$  and  $\theta = \frac{\pi}{10}$ ,  $20 \leq n \leq 40$ , suggest that *if the Green's function has negative values*, then the minimal value is obtained with the unit mass located at

$$P = P(n, k) = \frac{k}{n}(\mathbf{a} + \mathbf{b}),$$

(refer to Figure 4); more precisely

$$\mu(n, k) = \min_{x \in D} G_P^{n,k}(x).$$

Note that  $P$  is the point with the largest number of connections: 7. Thus we compute  $k(n)$  by looking only at  $G_P^{n,k}$ . Our numerical computations suggest that  $k(n)$  is bounded; more precisely there is a constant  $K = K(\theta)$  such that  $k(n) = K$  for  $n \geq N(\theta)$ . We got  $K(\frac{\pi}{10}) = 15$ ,  $N(\frac{\pi}{10}) = 63$  and  $K(\frac{\pi}{8}) = 10$ ,  $N(\frac{\pi}{8}) = 24$ , as can be read from Figure 5. In both cases we computed up to  $n = 300$ .

We restate the results of the experiment (take  $h = 1/n$ ): for the class of meshes considered, if  $h$  is small enough, the Green's function is nonnegative if only edges lying in a thin boundary layer of width  $Kh$  are nonnegative, with  $K$  depending on the chunkiness parameter of the mesh (see [1] or Section 5 for a precise definition).

We conclude this section with the following example.

**Example 3.4.** Let  $\Omega = [0, 1] \times [0, 1]$ , and let  $S$  be the linear (orientation preserving, to make a choice) transformation that takes  $D$  into  $\Omega$ . The mesh  $\mathcal{T}_{n,k}$  is mapped under  $S$  onto a mesh on the square denoted by  $\mathcal{U}_{n,k}$ , which has only nonnegative edges. Then  $G^n \circ S$  is the Green's function for the following operator on  $\Omega$ :

$$Lu = \nabla \cdot (A \nabla u)$$

with

$$A = \frac{1}{\sin^2 \theta} \begin{bmatrix} 1 & -\cos \theta \\ -\cos \theta & 1 \end{bmatrix}.$$

Notice that a priori the mesh  $\mathcal{U}_{n,k}$  is “nice” for all  $0 \leq k \leq n/2$  and that  $A$  is positive definite and well conditioned as long as  $\theta$  stays away from 0 (its eigenvalues are  $1 \pm \cos \theta$ ). However, if  $\theta < \pi/4$ , then Example 3.3 implies that whether the maximum principle is satisfied or not as  $n \rightarrow \infty$  depends strongly on  $k$ ; namely for  $k < K(\theta)$  it does not hold, while for  $k \geq K(\theta)$  it does, at least for the two  $\theta$ 's considered.

#### 4. THE DISCRETE MAXIMUM PRINCIPLE

In this section we discuss the conditions under which the discrete analogue of (1.2) follows from (1.10).

In the continuous homogeneous case there are three ingredients for proving (1.2) from (1.3):

- the integral representation of the solution  $u = U(g)$  yields

$$(4.1) \quad u(x) = - \int_{\partial D} \frac{\partial G_x}{\partial \nu}(\chi) g(\chi) d\chi \stackrel{\text{def}}{=} [U(g)](x);$$

- on the boundary we have

$$(4.2) \quad \frac{\partial G_x}{\partial \nu} < 0, \text{ on } \partial D;$$

- $U(1) = 1$ , and hence

$$(4.3) \quad - \int_{\partial D} \frac{\partial G_x}{\partial \nu} = 1, \forall x \in D.$$

It follows that the operator norm  $\mathcal{C}^0(\partial D) \xrightarrow{U} \mathcal{C}^0(D)$  is

$$(4.4) \quad \|U\| = \left( \sup_{x \in D} \int_{\partial D} \left| \frac{\partial G_x}{\partial \nu}(y) \right| dy \right) = 1$$

which implies (1.2).

We now turn our attention to the discrete homogeneous ( $f = 0$ ) problem (1.6). The following property is equivalent to the discrete version of (1.2):

$$(4.5) \quad \tilde{u}_h|_{\partial D} \geq 0 \Rightarrow u_h \geq 0 \text{ on } D, \text{ for all } \tilde{u}_h \in \mathcal{M}_h.$$

If we regard the boundary function as a column vector in  $\mathbf{B} \in \mathbb{R}^m$  with

$$\mathbf{B} = (b(P_{n+1}), \dots, b(P_{n+m}))^t,$$

then the right-hand side of the defining homogeneous linear system is (see also (1.9))  $\mathbf{r} = -\mathbf{H} \cdot \mathbf{B}$ , where  $\mathbf{H} \in \mathbb{R}^{n \times m}$ ,

$$(4.6) \quad \mathbf{H}_{ij} = a(\phi_i, \phi_{n+j}), \text{ for } 1 \leq i \leq n, 1 \leq j \leq m.$$

The discrete analogue of (4.1) is

$$(4.7) \quad \mathbf{U} = -\mathbf{A}^{-1} \mathbf{H} \mathbf{B}.$$

Since we want  $\mathbf{U} \geq 0$  for all  $\mathbf{B} \geq 0$ , it is necessary and sufficient that

$$(4.8) \quad \mathbf{A}^{-1} \mathbf{H} \leq 0.$$

Since  $\mathbf{A}^{-1} > 0$  is just a restatement of (1.10), a sufficient condition for (4.5) to hold is

$$(4.9) \quad \mathbf{H} \leq 0.$$

By Corollary A.2 in Appendix A, statement (4.9) is equivalent to the following condition on the mesh:

**Condition 4.1** (Boundary edge angle condition). Each edge with one vertex in the interior and the other one on the boundary is nonnegative.

We conclude this discussion with the following remark.

*Remark 4.2.* If the discrete Green’s function is nonnegative and the mesh satisfies Condition 4.1, then  $u_h$  satisfies the maximum principle (1.2) or its equivalent (4.5).

### 5. AN EXAMPLE OF A NEGATIVE GREEN’S FUNCTION

This section is organized as follows: in subsection 5.1 we describe the “bad” mesh and we state the main result, which we prove in subsection 5.2. In subsection 5.3 we show how one can use the “bad” mesh to violate the maximum principle (4.5).

**5.1. Preliminaries.** Let  $D$  be  $(-1, 1) \times (0, 1)$  with a “regular” triangulation  $\mathcal{T}_h$ , obtained by dividing the domain into equally sized squares of side  $h$  with sides parallel to the coordinate axes and by further cutting each square along its negative slope diagonal. We then further divide one triangle on the boundary  $\triangle NOP$  (see Figure 6), namely the one whose right angle is at the origin, in six triangles by adding three extra nodes,  $Q = Q(h, \epsilon) = (\frac{h}{4}, \epsilon h)$ ,  $R = R(h, \epsilon) = (\frac{3h}{4}, \epsilon h)$  inside, and  $S = S(h) = (\frac{h}{2}, 0)$  on the boundary (see Figure 7). We thus obtain the triangulation  $\mathcal{T}_{h,\epsilon}$ . Note that for  $\epsilon \ll 1$  the edge  $\overline{QR}$  is negative.

*Remark 5.1.* For fixed  $\epsilon$  the set of triangulations  $(\mathcal{T}_{h,\epsilon})_h$  is quasi-uniform.

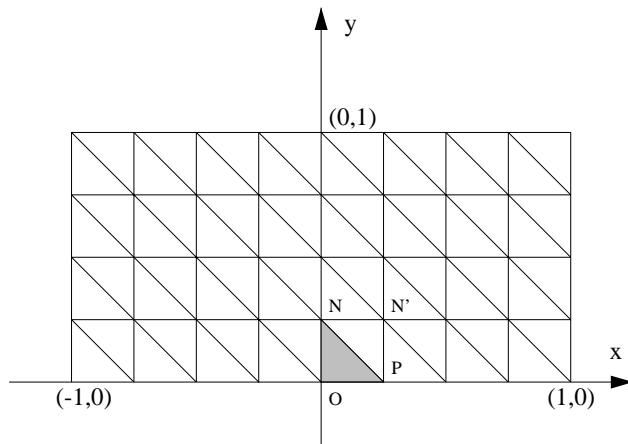


FIGURE 6. Mesh  $\mathcal{T}_h$  with  $h = \frac{1}{4}$ ; the filled triangle is further divided as in Figure 7 to obtain the mesh  $\mathcal{T}_{h,\epsilon}$ .

The **chunkiness**  $\gamma_T$  of a triangle  $T$  is defined in [1] (Definition (4.2.16)) to be

$$(5.1) \quad \gamma_T = \frac{\text{diam}(T)}{\sup\{\rho : T \text{ contains a ball of radius } \rho\}},$$

and similarly the chunkiness of a **triangulation**  $\mathcal{T}$  is

$$\gamma(\mathcal{T}) = \max\{\gamma_T : T \in \mathcal{T}\}.$$

Also we denote the (relative) diameter of a triangulation  $\mathcal{T}$  by

$$\text{diam}(\mathcal{T}) = \max\{\text{diam}(T) : T \in \mathcal{T}\} / \text{diam}(D).$$

For a quasi-uniform mesh  $\mathcal{T}$ , we associate  $h \approx \text{diam}(\mathcal{T})$ .

*Remark 5.2.* For fixed  $\epsilon$  the set  $(\gamma(\mathcal{T}_{h,\epsilon}))_h$  is bounded (the shape-regularity of the triangulations  $\mathcal{T}_{h,\epsilon}$  does not worsen with  $h \rightarrow 0$ ).

We denote by  $G_y^h$  (resp.  $G_y^{h,\epsilon}$ ) the discrete Green's function with singularity at  $y \in D$  associated with  $\mathcal{T}_h$  (resp.  $\mathcal{T}_{h,\epsilon}$ ).

We now state the main

**Theorem 5.3.** *There exists an  $\epsilon_0 > 0$  such that for all  $h > 0$  sufficiently small,*

$$(5.2) \quad G_R^{h,\epsilon_0}(Q) < 0.$$

*(That is, the Green's function corresponding to a unit mass source at  $R$  is negative at  $Q$ ).*

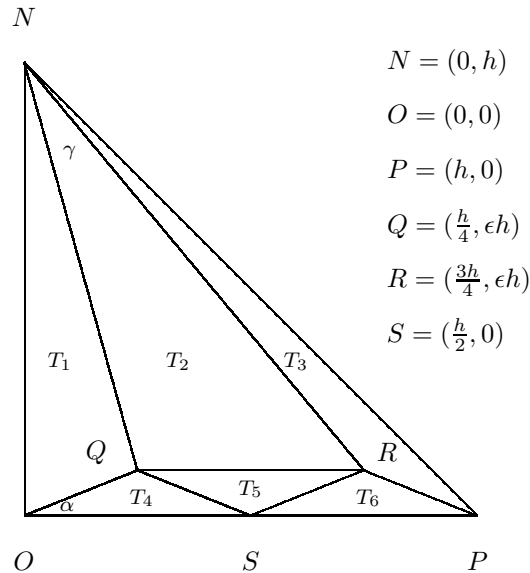


FIGURE 7. Divided triangle  $T_0$  near the boundary; the edge  $QR$  is negative.

Numerical computations show that  $\epsilon_0$  can be any number  $0 < \epsilon_0 < 0.0875$ . Note that the symmetry of the Green's function

$$(5.3) \quad G_R^{h,\epsilon_0}(Q) = G_Q^{h,\epsilon_0}(R)$$

implies also that  $G_Q^{h,\epsilon_0}(R) < 0$ .

Let us introduce some notation. Let  $(P_i)_{1 \leq i \leq n}$  be the set of *interior* vertices of  $\mathcal{T}_h$ , and we may assume  $P_1 = N$ . Also let  $\mathcal{B} = \{\phi_i\}_{1 \leq i \leq n} \subset \mathcal{M}_0(\mathcal{T}_h)$  be the corresponding nodal basis (for  $\mathcal{T}_h$ ). Let  $\mathcal{C} = \{\phi_Q, \phi_R\} \subset \mathcal{M}_0(\mathcal{T}_{h,\epsilon})$  be obtained by extending the nodal basis elements of  $T_0 = \triangle NOP$  to the whole domain  $D$  ( $T_0$  is triangulated as in Figure 7:  $T_0 = \bigcup_{i=1}^6 \{T_i\}$ ) by defining them to be 0 on  $\overline{D} \setminus T_0$ . Then  $\tilde{\mathcal{B}} = \mathcal{B} \cup \mathcal{C}$  is a basis for  $\mathcal{M}_0(\mathcal{T}_{h,\epsilon})$ , although *not* the standard nodal basis.

**5.2. Results.** The proof of Theorem 5.3 requires a few preliminary results.

*Remark 5.4.* If  $\psi \in W_1^1(T)$  and  $\phi : T \rightarrow \mathbb{R}$ ,  $\phi$  is linear and  $\psi(x) = 0$  for  $x \in \partial T$ , then  $\int_T \nabla \phi \cdot \nabla \psi = 0$  (integration by parts).

**Lemma 5.5** (*a-orthogonality*).  $\phi_Q$  and  $\phi_R$  are *a-orthogonal* to  $\mathcal{M}(\mathcal{T}_h)$ .

*Proof.* Recall that  $\text{supp}(\phi_Q), \text{supp}(\phi_R) \subset T_0$ ; hence  $\phi_Q$  and  $\phi_R$  vanish on  $\partial T_0$ . On the other hand any function in  $\mathcal{M}(\mathcal{T}_h)$  is linear on  $T_0$ , since  $T_0$  is part of the triangulation  $\mathcal{T}_h$ . The conclusion follows from Remark 5.4.  $\square$

We denote by  $\mathbf{a}$  the  $2 \times 2$  stiffness matrix corresponding to  $\phi_Q$  and  $\phi_R$ , i.e.,

$$(5.4) \quad \mathbf{a}_{\alpha,\beta} = a(\phi_\beta, \phi_\alpha), \text{ for } \alpha, \beta = Q, R.$$

**Lemma 5.6.** *The inverse of  $\mathbf{a}$  has the form*

$$(5.5) \quad \mathbf{a}^{-1} = \mathbf{a}^{-1}(\epsilon) = \frac{\epsilon}{\delta} \begin{bmatrix} M & -L \\ -L & K \end{bmatrix}$$

with  $M(\epsilon), K(\epsilon) \rightarrow_{\epsilon \rightarrow 0} 5$ ,  $L(\epsilon) \rightarrow_{\epsilon \rightarrow 0} 1$ , and  $\delta(\epsilon) \rightarrow_{\epsilon \rightarrow 0} 24$ .

We give the proof of Lemma 5.6 in Appendix B.

The following result is essential to the proof of Theorem 5.3. It states that there is a boundary layer of width  $O(h)$  on which the Green's function is bounded. It will be applied **only to  $\mathcal{T}_h$** . Since this surprising result is of independent interest, we state it as

**Theorem 5.7** (A bound for the value of the Green's function). *Given a domain  $D \subset \mathbb{R}^2$  and a quasi-uniform set of triangulations  $(\mathcal{T}_h)_h$  ( $\text{diam}(\mathcal{T}_h) \leq h$ ) of  $D$  with bounded chunkiness (for all  $h > 0$ ,  $\gamma(\mathcal{T}_h) \leq \gamma$ ), and a number  $K > 0$ , there exists a constant  $C = C(K, \gamma)$  such that for all  $x, y \in D$  both within distance  $Kh$  from  $\partial D$  we have*

$$(5.6) \quad \|G_x^h\|_{H^1} \leq C$$

and

$$(5.7) \quad |G_x^h(y)| \leq C.$$

In particular

$$(5.8) \quad |G_x^h(x)| \leq C.$$

*Proof.* As usual we denote by  $|\cdot|_{H^1}$  (resp.  $\|\cdot\|_{H^1}$ ) the  $H^1$  seminorm (resp. norm). In what follows  $C$  stands for a generic constant independent of  $h$ . Recall that  $G_x^h$  satisfies

$$(5.9) \quad \int_D \nabla G_x^h \cdot \nabla v = v(x) \quad \forall v \in \mathcal{M}_{0,h}.$$

By taking  $v = G_x^h$ , we get

$$(5.10) \quad |G_x^h|_{H^1(D)}^2 = G_x^h(x).$$

Let  $\eta$  be a segment connecting  $x$  with  $x_0 \in \partial D$  with  $\text{length}(\eta) \leq Kh$ . The mean value theorem on  $\eta$  together with  $G_x^h(x_0) = 0$  implies

$$(5.11) \quad G_x^h(x) \leq \text{length}(\eta) \cdot |G_x^h|_{W_\infty^1(D)} \leq Kh |G_x^h|_{W_\infty^1(D)}.$$

Inverse estimates (Theorem 4.5.11 in [1]) imply

$$(5.12) \quad \|G_x^h\|_{W_\infty^1(D)} \leq Ch^{-1} \|G_x^h\|_{H^1(D)}$$

since the set  $(\mathcal{T}_h)_h$  is quasi-uniform. Putting together (5.10), (5.11), (5.12) and the fact that  $\|\cdot\|_{H^1(D)} \leq C|\cdot|_{H^1(D)}$  holds for  $H^1$  functions that are 0 on  $\partial D$ , we get

$$\begin{aligned} \|G_x^h\|_{H^1(D)}^2 &\leq C|G_x^h|_{H^1(D)}^2 \\ &= CG_x^h(x) \\ &\leq CKh |G_x^h|_{W_\infty^1(D)} \\ &\leq CKhh^{-1} \|G_x^h\|_{H^1(D)} = CK \|G_x^h\|_{H^1(D)}, \end{aligned}$$

which implies

$$(5.13) \quad \|G_x^h\|_{H^1(D)} \leq CK.$$

Now by taking  $v = G_y^h$  in (5.9), we get

$$(5.14) \quad G_y^h(x) = \int_D \nabla G_x^h \cdot \nabla G_y^h;$$

hence

$$(5.15) \quad |G_x^h(y)| = |G_y^h(x)| \leq |G_x^h|_{H^1(D)} |G_y^h|_{H^1(D)} \stackrel{\text{by (5.13)}}{\leq} (CK)^2.$$

□

*Remark 5.8.* To make the argument in Theorem 5.7 work in  $d \geq 3$  dimensions, we would need to change the scaling  $d(x, \partial D) < Kh$  to

$$(5.16) \quad d(x, \partial D) < Kh^{\frac{d}{2}}$$

in order to accommodate the  $d$ -dimensional equivalent of (5.12), which is

$$(5.17) \quad \|G_x^h\|_{W_\infty^1(D)} \leq Ch^{-d/2} \|G_x^h\|_{H^1(D)}.$$

However, no interior vertices are expected to satisfy (5.16) for  $d \geq 3$  and small  $h$  since for shape-regular simplices the interior vertices are at least  $O(h)$  away from  $\partial D$ . This suggests that if  $x$  is a mesh node, then  $G_x^h(x)$  grows infinitely large as  $h \rightarrow 0$  in  $d \geq 3$  dimensions. In fact we expect that in this case

$$(5.18) \quad G_x^h(x) = O(h^{2-d}) \text{ for all } x \text{ (up to the boundary).}$$

We now proceed to the proof of Theorem 5.3.

*Proof.* By Lemma 5.5,  $\phi_Q$  and  $\phi_R$  are  $\alpha$ -orthogonal to all basis functions in  $\mathcal{B}$ . Hence the stiffness matrix  $\tilde{\mathbf{A}}$  written in the basis  $\tilde{\mathcal{B}}$  decouples as

$$(5.19) \quad \tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{a} \end{bmatrix},$$

where  $\mathbf{A}$  is the stiffness matrix corresponding to  $\mathcal{T}_h$ . For  $f \in \mathcal{M}_h^*$  equation (1.6) translates into two systems:

$$(5.20) \quad \mathbf{A}\boldsymbol{\kappa} = \mathbf{b}$$

$$(5.21) \quad \mathbf{a}\boldsymbol{\lambda} = \mathbf{c},$$

with  $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_n)^t \in \mathbb{R}^n$ ,  $\boldsymbol{\lambda} = (\lambda_Q, \lambda_R)^t \in \mathbb{R}^2$  and

$$(5.22) \quad b_i = \langle f, \phi_i \rangle, \text{ for } 1 \leq i \leq n,$$

$$(5.23) \quad c_\alpha = \langle f, \phi_\alpha \rangle, \text{ for } \alpha = Q, R.$$

Accordingly the solution  $u_h$  can be written as

$$(5.24) \quad u_h = \underbrace{\sum_{i=1}^n \kappa_i \phi_i}_{u_h^{(1)}} + \underbrace{\lambda_Q \phi_Q + \lambda_R \phi_R}_{u_h^{(2)}}.$$

Equation (5.20) implies  $\boldsymbol{\kappa} = \sum_{i=1}^n b_i \mathbf{A}^{-1} \mathbf{e}_i$ , where  $\mathbf{e}_i$  is the  $i^{th}$  column of the identity matrix of  $\mathbb{R}^n$ , and we prefer to regard  $u_h^{(1)}$  as

$$(5.25) \quad u_h^{(1)} = \sum_{i=1}^n b_i G_{P_i}^h,$$

with  $G_{P_i}^h$  being the discrete Green's function **associated with the mesh  $\mathcal{T}_h$  at  $P_i$** , as before.

We now take  $f$  to be the unit measure at  $R$  ( $f = \delta_R$ ). Then

$$\begin{aligned} b_1 &= \phi_N(R) = \epsilon, \\ b_i &= 0 \text{ for } 2 \leq i \leq n, \\ c_Q &= \phi_Q(R) = 0, \\ c_R &= \phi_R(R) = 1. \end{aligned}$$

Hence the solution can be written as

$$(5.26) \quad G_R^{h,\epsilon} = \epsilon G_N^h + \lambda_Q \phi_Q + \lambda_R \phi_R.$$

Its value at  $Q$  is

$$(5.27) \quad G_R^{h,\epsilon}(Q) = \epsilon G_N^h(Q) + \lambda_Q = \epsilon^2 G_N^h(N) + \lambda_Q.$$

Lemma 5.6 and (5.21) imply that

$$(5.28) \quad \lambda_Q = -\frac{\epsilon L}{\delta} \approx_{\epsilon \rightarrow 0} -\frac{\epsilon}{24}.$$

Hence

$$(5.29) \quad G_R^{h,\epsilon}(Q) = \epsilon \left( \epsilon G_N^h(N) - \frac{L}{\delta} \right) \approx_{\epsilon \rightarrow 0} \epsilon \left( \epsilon G_N^h(N) - \frac{1}{24} \right).$$

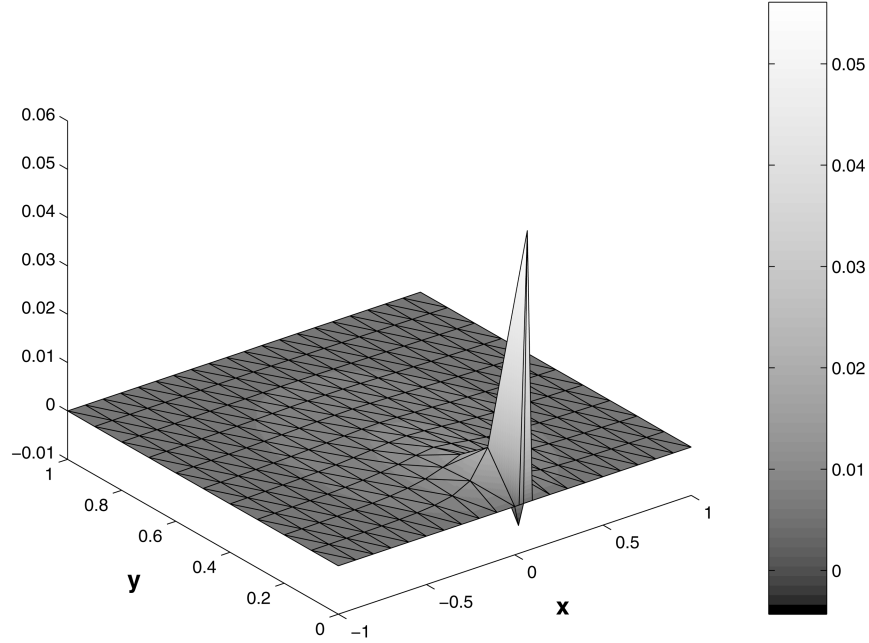


FIGURE 8.  $G_R^{h, \epsilon_0}$  with  $h = 10^{-1}$ ,  $\epsilon_0 = 2.5 \times 10^{-2}$

Recall that by Theorem 5.7,  $G_N^h(N)$  is bounded by a constant  $C$  as  $h \rightarrow 0$ , and there exists  $\epsilon_1 > 0$  such that

$$(5.30) \quad \epsilon < \frac{L(\epsilon)}{\delta(\epsilon)C}, \quad \forall \epsilon < \epsilon_1.$$

This implies

$$(5.31) \quad \epsilon G_N^h(N) < \epsilon C < \frac{L(\epsilon)}{\delta(\epsilon)}, \quad \forall \epsilon < \epsilon_1.$$

Thus  $G_R^{h, \epsilon}(Q) < 0$ . Hence  $\epsilon_0$  can be any number in  $(0, \epsilon_1)$ .  $\square$

Numerical simulations performed with Matlab suggest that  $G_N^h(N)$  **increases** as  $h \searrow 0$  (we can easily show  $G_N^{2h}(N) < G_N^h(N)$ ) and converges to a value  $G_N^{0+}(N) \approx 0.3634$ ; the largest value for  $\epsilon_1$ , i.e., where  $\epsilon_1 G_N^{0+}(N) = L(\epsilon_1)/\delta(\epsilon_1)$ , is approximately 0.0875. Figure 8 shows the graph of  $G_R^{h, \epsilon}$  for  $h = 10^{-1}$ ,  $\epsilon = 2.5 \times 10^{-2}$ .

We would like to remark on the size of the negative part of  $G_R^{h, \epsilon}$ . As expected and confirmed by numerical computations, the Green's function  $G_R^{h, \epsilon}$  attains its maximum at  $R$ . By (5.26) we get

$$(5.32) \quad G_R^{h, \epsilon}(R) = \epsilon^2 G_N^h(N) + \lambda_R.$$

From (5.21) and (5.5) we obtain

$$(5.33) \quad \lambda_R = \epsilon \frac{K(\epsilon)}{\delta(\epsilon)} \approx_{\epsilon \rightarrow 0} \frac{5\epsilon}{24},$$

which implies that for a small but fixed  $\epsilon_0 > 0$  the negative part of  $G_R^{h,\epsilon_0}$  is significant in size (about 1/5) compared to its positive part as  $h \rightarrow 0$ . This calculation also points to the observed fact that

$$(5.34) \quad \|G_R^{h,\epsilon}\|_{L^\infty} = O(\epsilon).$$

**5.3. A discrete harmonic function, nonnegative on the boundary and with nonzero negative part.** Our next calculation shows that despite the smallness of  $G_R^{h,\epsilon}$  as shown in (5.34), its relatively large negative part can influence (4.5) to fail significantly, even if we let  $\epsilon \rightarrow 0$ .

Let us define  $\phi_P \in \mathcal{M}(\mathcal{T}_{h,\epsilon})$  to be the nodal basis element associated with  $P$  (i.e.,  $\phi_P(P) = 1, \phi_P(V) = 0$  for all other vertices  $V$  of  $\mathcal{T}_{h,\epsilon}$  (this corresponds to the hat function  $g \in \mathcal{C}^0(\partial D)$  over  $P$ ; refer to Figure 7).

We compute the value  $u_h(Q)$ , where  $u_h$  solves (1.6) with  $\tilde{u}_h = \phi_P$ . Let  $N' \in \mathbb{R}^2$  denote the point  $(h, h)$  (see Figure 7) and assume  $N'$  has index 2 (i.e.,  $P_2 = N'$  in the previous notation). The only nonzeros in the right-hand sides of the systems (5.20) and (5.21) are

$$(5.35) \quad b_1 = -a(\phi_P, \phi_N),$$

$$(5.36) \quad b_2 = -a(\phi_P, \phi_{N'}) = -1,$$

$$(5.37) \quad c_R = -a(\phi_P, \phi_R),$$

since  $\phi_N, \phi_{N'}$  and  $\phi_R$  are the only ones whose support intersects  $\text{supp}(\phi_P)$ . Equation (5.25) implies

$$(5.38) \quad u_h = \phi_P + b_1 G_N^h + b_2 G_{N'}^h + \lambda_Q \phi_Q + \lambda_R \phi_R,$$

where (as before)

$$(5.39) \quad \begin{bmatrix} \lambda_Q \\ \lambda_R \end{bmatrix} = \mathbf{a}^{-1} \begin{bmatrix} 0 \\ c_R \end{bmatrix}$$

$$(5.40) \quad = \frac{\epsilon}{\delta} \begin{bmatrix} M & -L \\ -L & K \end{bmatrix} \begin{bmatrix} 0 \\ c_R \end{bmatrix}$$

$$(5.41) \quad = \frac{\epsilon}{\delta} \begin{bmatrix} -L c_R \\ K c_R \end{bmatrix}.$$

We now compute  $c_R$  and  $b_1$ . If we denote by  $\zeta = \zeta(\epsilon)$  the angle  $\angle RNP$ , then by formula (A.2)

$$(5.42) \quad c_R = -a(\phi_P, \phi_R) = \frac{\sin(\alpha + \zeta)}{2 \sin \alpha \sin \zeta} = \frac{2J(\epsilon)}{\epsilon},$$

with  $\lim_{\epsilon \rightarrow 0} J(\epsilon) = \lim_{\epsilon \rightarrow 0} \frac{\sin(\alpha + \zeta)}{\sin \zeta} = 1$  (recall that  $\alpha \approx \epsilon/4$  for  $\epsilon \ll 1$ ). Hence

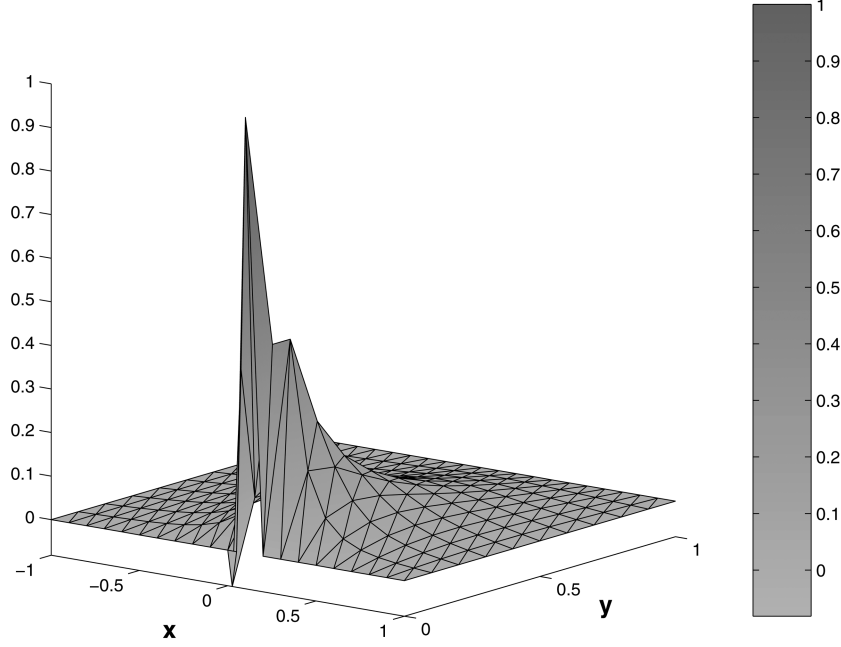
$$(5.43) \quad \lim_{\epsilon \rightarrow 0} \lambda_Q = - \lim_{\epsilon \rightarrow 0} \frac{2J(\epsilon)}{\epsilon} \frac{\epsilon}{\delta} L(\epsilon) = - \lim_{\epsilon \rightarrow 0} \frac{2J(\epsilon)}{\delta} L(\epsilon) = -\frac{1}{12}.$$

Since  $\text{supp}(\phi_P) \cap \text{supp}(\phi_N) = T_3 \cup T_6$  (remember  $\phi_N$  lives on  $\mathcal{T}_h$ ), we get

$$(5.44) \quad b_1 = -a(\phi_P, \phi_N) = - \int_{T_3 \cup T_6} \nabla \phi_P \cdot \nabla \phi_N.$$

One can see that  $\triangle NRP$  converges to a nondegenerate triangle as  $\epsilon \rightarrow 0$ ; hence

$$(5.45) \quad \int_{T_3} \nabla \phi_P \cdot \nabla \phi_N = J_1(\epsilon),$$

FIGURE 9.  $u_h$  with  $h = 10^{-1}$ ,  $\epsilon_0 = 10^{-3}$ 

with  $J_1(\epsilon)$  converging to a finite positive value. On  $T_6$  the angle  $\theta = \angle(\nabla\phi_P, \vec{OY})$  becomes nearly  $\pi$  as  $\epsilon \rightarrow 0$ ; hence

$$\begin{aligned}
 \int_{T_6} \nabla\phi_P \cdot \nabla\phi_N &= \frac{1}{h} \int_{T_6} Pr_Y \nabla\phi_N \\
 &= \frac{\cos(\theta)}{h} \text{area}(T_6) |\phi_N| \\
 &= \frac{\cos(\theta)}{h} \frac{h}{4} \frac{h}{4} \tan \alpha \left( \frac{h}{2} \sin \alpha \right)^{-1} \\
 &= \cos(\theta) \frac{1}{16} \tan \alpha \left( \frac{1}{2} \sin \alpha \right)^{-1} \\
 &= \frac{\cos \theta}{8 \cos \alpha} \rightarrow_{\epsilon \rightarrow 0} -\frac{1}{8},
 \end{aligned}$$

where  $Pr_Y \vec{v}$  is simply the orthogonal projection in  $\mathbb{R}^2$  of the vector  $\vec{v} \in \mathbb{R}^2$  onto the  $Y$  axis. It follows that

$$(5.46) \quad b_1 = -a(\phi_P, \phi_N) \approx_{\epsilon \rightarrow 0} J_1(\epsilon) - \frac{1}{8};$$

hence  $\lim_{\epsilon \rightarrow 0} b_1(\epsilon)$  is finite. By (5.38)

$$(5.47) \quad u_h(Q) = b_1(\epsilon) G_N^h(Q) + b_2 G_{N'}^h(Q) + \lambda_Q(\epsilon)$$

$$(5.48) \quad = \epsilon (b_1(\epsilon) G_N^h(N) + G_{N'}^h(N)) + \lambda_Q(\epsilon).$$

Theorem 5.7 implies that  $G_N^h(N)$  and  $G_{N'}^h(N)$  are bounded as  $h \rightarrow 0$ ; therefore for a fixed small enough  $\epsilon_0$  and small enough  $h$

$$(5.49) \quad u_h(Q) \approx -\frac{1}{12}.$$

It follows that the negative part of  $u_h$  is  $O(1)$ . In Figure 9, produced using Matlab’s “pdetoolbox”, we show the graph of  $u_h$  for  $h = 10^{-1}$  and  $\epsilon = 10^{-3}$ . It should be noted that  $u_h$  has a local maximum in the interior (at  $N'$ ); this is not surprising since the presence of a negative edge implies that  $u_h$  is not a *convex combination function* (see [3] for details). Furthermore the value  $u_h(Q)$  is approximately  $-0.0815 \approx -1/12$ .

### 6. MONOTONICITY PRESERVATION FOR A CLASS OF MESH OPERATIONS

Throughout this section we call a mesh *monotone* (resp. *strictly monotone*) if the corresponding discrete Green’s function obtained from (1.6) is nonnegative (resp. positive). It might appear from the preceding sections that it is easy to construct nonmonotone meshes, but now we show that in fact it is relatively difficult to do so. Moreover, the same techniques can be used.

The computation needed in the proof of Theorem 5.3 was facilitated by the decoupling of the stiffness matrix as shown in (5.19). The same idea leads to the following theorem.

**Theorem 6.1.** *Let  $\mathcal{T}_h$  be a monotone (resp. strictly monotone) mesh on  $D$ , and let  $\mathcal{T}_{h,1}$  be the mesh obtained by cutting one mesh-triangle  $T = \triangle PQR$  (of  $\mathcal{T}_h$ ) along the lines  $NP, NQ$  and  $NR$ , where  $N \in \text{Int}(T)$  (see upper triangle in Figure 10). Then  $\mathcal{T}_{h,1}$  is monotone (resp. strictly monotone).*

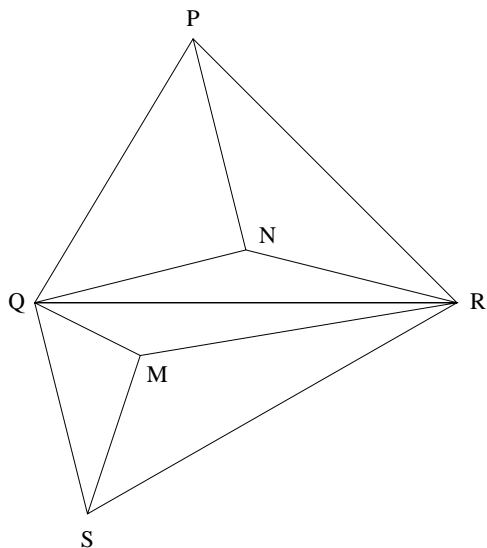


FIGURE 10. A mesh obtained by adding the points  $M$  and  $N$  to a monotone mesh remains monotone, even though the edge  $QR$  is negative.

*Proof.* As in the proof of Theorem 5.3 we use a hierarchical basis on  $\mathcal{M}_0(\mathcal{T}_{h,1})$ , namely  $\tilde{\mathcal{B}} = \mathcal{B} \cup \{\phi_N\}$ , where  $\mathcal{B}$  is the nodal basis for  $\mathcal{M}_0(\mathcal{T}_h)$ ,  $\phi_N \in \mathcal{M}_0(\mathcal{T}_{h,1})$  vanishes on all vertices of  $\mathcal{T}_h$  and  $\phi_N(N) = 1$ . By Remark 5.4,  $a(\phi_N, \phi) = 0$  for all  $\phi \in \tilde{\mathcal{B}}$ , which implies that the stiffness matrix  $\tilde{\mathbf{A}}$  written in the basis  $\tilde{\mathcal{B}}$  decouples as

$$(6.1) \quad \tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & d \end{bmatrix},$$

where  $\mathbf{A}$  is the stiffness matrix corresponding to  $\mathcal{T}_h$  and  $d = a(\phi_N, \phi_N) > 0$ . For  $x \in D$  we write  $G_x^{h,1} = \sum_i \lambda_i \phi_i + \lambda_N \phi_N$ , where  $\tilde{\boldsymbol{\lambda}} = \underbrace{(\lambda_1, \dots, \lambda_n)}_{\boldsymbol{\lambda}^\dagger}, \lambda_N)^\dagger$  is the solution

of

$$(6.2) \quad \tilde{\mathbf{A}}\tilde{\boldsymbol{\lambda}} = \tilde{\mathbf{b}}$$

with  $\tilde{\mathbf{b}} = \underbrace{(b_1, \dots, b_n)}_{\mathbf{b}^\dagger}, b_N)^\dagger$  defined by  $b_i = \phi_i(x), b_N = \phi_N(x)$ . Note that for any  $x$  at least one of the first  $n$   $b_i$ 's is positive, and there can be no more than four of them positive and all are nonnegative. Equation (6.2) is equivalent to

$$(6.3) \quad \boldsymbol{\lambda} = \mathbf{A}^{-1}\mathbf{b},$$

$$(6.4) \quad \lambda_N = b_N/d;$$

hence  $\lambda_i = G_x^h(P_i)$  for  $1 \leq i \leq n$ . Therefore

$$(6.5) \quad G_x^{h,1} = \sum_i G_x^h(P_i)\phi_i + \frac{b_N}{d}\phi_N.$$

This is a nonnegative combination of nonnegative functions; hence  $G_x^{h,1}$  is nonnegative. If  $G^h$  is positive, then *all* coefficients  $G_x^h(P_i)$  in (6.5) are positive, which makes  $G_x^{h,1}$  positive.  $\square$

The following corollary, although evident, shows how one can construct monotone meshes which contain many negative edges (see also Figure 10).

**Corollary 6.2.** *A mesh obtained by inductively cutting triangles as described in Theorem 6.1 starting with a monotone mesh is monotone.*

The extension of the results in this section to higher dimensions is trivial.

#### APPENDIX A. TRIGONOMETRIC FORMULAS FOR THE STIFFNESS MATRIX

Let  $\mathbf{e} = AB$  be an edge bordered by two triangles  $T_1 = \triangle ABC$  and  $T_2 = \triangle ABD$  in  $\mathcal{T}_h$ . Let  $A' \in BC$  and  $B' \in AC$  be such that  $A'A \perp BC$  and  $B'B \perp AC$ , and let  $h_A = \|A'A\|$  and  $h_B = \|B'B\|$ . Denote by  $\alpha$  (resp.  $\beta, \gamma, \delta$ ) the angle  $\angle BAC$  (resp.  $\angle ABC, \angle ACB, \angle ADB$ ) as described in Figure 11, and denote by  $\zeta$  the angle between the vectors  $A - A'$  and  $B - B'$ . Furthermore, let  $\sigma$  be the area of  $T = \triangle ABC$ ,  $a = \|BC\|$ ,  $b = \|CA\|$ ,  $c = \|AB\|$ , and recall that  $\phi_A$  (resp.  $\phi_B$ ) is the nodal basis function at  $A$  (resp.  $B$ ). Since  $\phi_A$  is linear on  $T_1$ , it follows that

$$\nabla\phi_A|_{T_1} \equiv \frac{A - A'}{h_A^2}$$

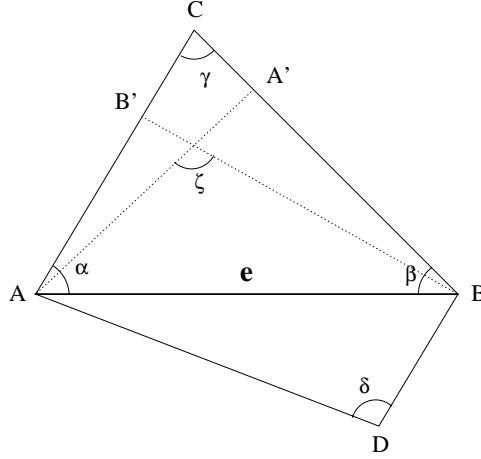


FIGURE 11.

and a similar formula holds for  $\phi_B$ . Then

$$\begin{aligned} \int_{T_1} \nabla \phi_A \cdot \nabla \phi_B &= \int_T \frac{A - A'}{h_A^2} \cdot \frac{B - B'}{h_B^2} dx = \sigma \frac{\cos \zeta}{h_A h_B} = \frac{1}{2} a h_A \frac{\cos \zeta}{h_A h_B} \\ &= \frac{1}{2} a h_A \frac{\cos(\pi - \gamma)}{h_A h_B} = -\frac{\cos \gamma}{2} \frac{a}{h_B} \\ &= -\frac{\cos \gamma}{2} \frac{a}{c \sin \alpha} = -\frac{\cos \gamma}{2} \frac{\sin \alpha}{\sin \gamma \sin \alpha} = -\frac{\cot \gamma}{2}. \end{aligned}$$

We have used  $\zeta = \pi - \gamma$  and  $h_B = c \sin \alpha = a \sin \gamma$ . Now

$$\int_{T_1} \|\nabla \phi_A\|^2 = \frac{\sigma}{h_A^2} = \frac{a}{2h_A} = \frac{a}{2c \sin \beta} = \frac{\sin \alpha}{2 \sin \gamma \sin \beta}.$$

We summarize these formulas in the following lemma.

**Lemma A.1.** *With the notation in Figure 11 we have*

$$(A.1) \quad \int_{T_1} \|\nabla \phi_A\|^2 = \frac{\sin \alpha}{2 \sin \beta \sin \gamma},$$

$$(A.2) \quad \int_{T_1 \cup T_2} \nabla \phi_A \cdot \nabla \phi_B = -\frac{1}{2} (\cot \gamma + \cot \delta) = -\frac{\sin(\gamma + \delta)}{2 \sin \gamma \sin \delta}.$$

A simple consequence of formula (A.2) is the following.

**Corollary A.2.** *With the notation from the previous lemma*

$$(A.3) \quad a(\phi_A, \phi_B) \leq 0 \text{ iff } \gamma + \delta \leq \pi.$$

## APPENDIX B. PROOF OF LEMMA 5.6

*Proof.* We use formulas (A.1) and (A.2) in Appendix A to compute the entries in the stiffness matrix  $\mathbf{a}$  given by the triangles  $T_1, \dots, T_6$  (refer to Figure 7). One can easily see that as  $\epsilon \rightarrow 0$  all angles outside the quadrilateral  $OQRP$  stay bounded

away from 0 or  $\pi$ ; hence their trigonometric functions are bounded and bounded away from 0. If we define  $\gamma = \angle QNR$ , then

$$(B.1) \quad \begin{aligned} a(\phi_Q, \phi_R) &= -\frac{\sin(\pi - 2\alpha + \gamma)}{2 \sin(\pi - 2\alpha) \sin \gamma} \\ &= -\frac{\sin(\pi - 2\alpha + \gamma)}{4 \sin \alpha \cos \alpha \sin \gamma}. \end{aligned}$$

Since  $\alpha(\epsilon) \approx \epsilon/4$  for  $\epsilon \rightarrow 0$  and  $\sin(\pi - 2\alpha + \gamma) < 0$ , equation (B.1) becomes

$$(B.2) \quad a(\phi_Q, \phi_R) = \frac{L}{\epsilon},$$

where  $L = L(\epsilon)$  and  $\lim_{\epsilon \rightarrow 0} L(\epsilon) = -\frac{\sin(\pi + \gamma_0)}{\sin \gamma_0} = 1$ , where  $\gamma_0 = \lim_{\epsilon \rightarrow 0} \gamma(\epsilon) \in (0, \pi/4)$  (see Figure 7). We now turn our attention to the diagonal entries:

$$(B.3) \quad a(\phi_Q, \phi_Q) = \sum_{i=1,2,4,5} \int_{T_i} |\nabla \phi_Q|^2$$

$$(B.4) \quad = K_1(\epsilon) + \int_{T_4} |\nabla \phi_Q|^2 + \int_{T_5} |\nabla \phi_Q|^2$$

$$(B.5) \quad = K_1(\epsilon) + \frac{\sin(\pi - 2\alpha)}{2 \sin^2 \alpha} + \frac{\sin \alpha}{2 \sin(\pi - 2\alpha) \sin \alpha}$$

$$(B.6) \quad = K_1(\epsilon) + \frac{\cos \alpha}{\sin \alpha} + \frac{1}{4 \sin \alpha \cos \alpha}$$

$$(B.7) \quad = \frac{1}{4 \sin \alpha} \left( K_1(\epsilon) \sin \alpha + 4 \cos \alpha + \frac{1}{\cos \alpha} \right).$$

Hence

$$(B.8) \quad a(\phi_Q, \phi_Q) = \frac{K(\epsilon)}{\epsilon},$$

where  $\lim_{\epsilon \rightarrow 0} K(\epsilon) = 5$ . Similarly

$$(B.9) \quad a(\phi_R, \phi_R) = \frac{M(\epsilon)}{\epsilon}$$

where  $\lim_{\epsilon \rightarrow 0} M(\epsilon) = 5$ . The matrix  $\mathbf{a}$  takes the form  $\mathbf{a} = \frac{1}{\epsilon} \mathbf{a}'$ ,

$$(B.10) \quad \mathbf{a}' = \mathbf{a}'(\epsilon) = \begin{bmatrix} K & L \\ L & M \end{bmatrix}$$

and

$$(B.11) \quad \lim_{\epsilon \rightarrow 0} \mathbf{a}' = \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}.$$

It follows that

$$(B.12) \quad \mathbf{a}^{-1} = \frac{\epsilon}{\det(\mathbf{a}')} \begin{bmatrix} M & -L \\ -L & K \end{bmatrix},$$

and  $\delta = \det(\mathbf{a}') \rightarrow 24$ . □

#### ACKNOWLEDGMENTS

The authors thank Jim Bramble for helpful discussions and the anonymous editor for useful pointers to references.

## REFERENCES

1. Susanne C. Brenner and L. Ridgway Scott, *The mathematical theory of finite element methods*, second ed., Texts in Applied Mathematics, vol. 15, Springer-Verlag, New York, 2002. MR **2003a**:65103
2. P. G. Ciarlet and P.-A. Raviart, *Maximum principle and uniform convergence for the finite element method*, *Comput. Methods Appl. Mech. Engrg.* **2** (1973), 17–31. MR **51**:11992
3. Michael S. Floater, *One-to-one piecewise linear mappings over triangulations*, *Math. Comp.* **72** (2003), no. 242, 685–696 (electronic).
4. W. Höhn and H.-D. Mittelmann, *Some remarks on the discrete maximum-principle for finite elements of higher order*, *Computing* **27** (1981), no. 2, 145–154. MR **83a**:65109
5. Sergey Korotov, Michal Krížek, and Pekka Neittaanmäki, *Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle*, *Math. Comp.* **70** (2001), no. 233, 107–119 (electronic). MR **2001i**:65126
6. Kurt Kreith, *Criteria for positive Green's functions*, *Illinois J. Math.* **12** (1968), 475–478. MR **37**:4400
7. Vitoriano Ruas Santos, *On the strong maximum principle for some piecewise linear finite element approximate problems of nonpositive type*, *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **29** (1982), no. 2, 473–491. MR **84b**:65118
8. Alfred H. Schatz, *A weak discrete maximum principle and stability of the finite element method in  $L_\infty$  on plane polygonal domains. I*, *Math. Comp.* **34** (1980), no. 149, 77–91. MR **81e**:65063
9. Ridgway Scott, *Optimal  $L^\infty$  estimates for the finite element method on irregular meshes*, *Math. Comp.* **30** (1976), no. 136, 681–697. MR **55**:9560
10. Gilbert Strang and George J. Fix, *An analysis of the finite element method*, Prentice-Hall Inc., Englewood Cliffs, N. J., 1973, Prentice-Hall Series in Automatic Computation. MR **56**:1747
11. Richard S. Varga, *Matrix iterative analysis*, expanded ed., Springer Series in Computational Mathematics, vol. 27, Springer-Verlag, Berlin, 2000. MR **2001g**:65002
12. Jinchao Xu and Ludmil Zikatanov, *A monotone finite element scheme for convection-diffusion equations*, *Math. Comp.* **68** (1999), no. 228, 1429–1446. MR **99m**:65225

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CHICAGO, CHICAGO, ILLINOIS 60637  
E-mail address: [draga@cs.uchicago.edu](mailto:draga@cs.uchicago.edu)

DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF CHICAGO, CHICAGO, ILLINOIS 60637  
E-mail address: [t-dupont@uchicago.edu](mailto:t-dupont@uchicago.edu)

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CHICAGO, CHICAGO, ILLINOIS 60637  
E-mail address: [ridg@uchicago.edu](mailto:ridg@uchicago.edu)