Analysis of Domain Decomposition Preconditioners
on L-Shaped and C-Shaped Regions

Tony F. Chan
Diana C. Resasco

August 1988
CAM Report 88-29
We present an algebraic analysis of some domain decomposition preconditioners on irregular regions. We analyze a preconditioner proposed in [3] for the interface system and prove that, for all L-shaped regions and some C-shaped regions, it produces a convergence rate that is independent of the size of the discretization and the relative shape of the subdomains (aspect ratios). Specifically, we prove that the condition number of the preconditioned capacitance system is bounded by 2.16 for all L-shaped domains. We also give some results for other simple irregular geometries.

Analysis of Domain Decomposition Preconditioners on L-shaped and C-shaped Regions

Tony F. Chan † and Diana C. Resasco ‡
Research Report YALEU/DCS/RR-534
August 1988

The authors were supported in part by the Department of Energy under contract DE-FG03-87ER25037 at UCLA and DE-AC02-81ER10996 at Yale and also, by NSF under contract DMS-8714612 at UCLA and by the Army Research Office under contract DAAL 03/88/K/0085. Approved for public release: distribution is unlimited.

Keywords: domain decomposition, substructuring, capacitance matrix, irregular regions, elliptic partial differential equations.

†Dept. of Mathematics, UCLA, Los Angeles, CA 90024
‡Dept. of Computer Science, Research Center for Scientific Computing, Yale University, Box 2158, Yale Station, New Haven, Conn. 06520.
1. Introduction

We consider the problem of solving an elliptic partial differential equation on a domain that is broken up into rectangular subregions. By using domain decomposition or substructuring techniques, the problem is reduced to separately solving approximate problems in the subdomains and updating the solution at the interfaces between two or more subregions. For the class of domain decomposition methods considered in this paper, the basic idea consists of the following: the differential operator is discretized on a grid imposed over the domain, which is partitioned into several subregions. Then, by applying block elimination to the discretized equations, a system is derived for the unknowns on the interfaces between subregions. This system is sometimes called the capacitance system. Forming the right hand side for the interface system requires the solution of independent elliptic problems on the subdomains. For certain constant coefficient problems on regular domains, fast direct methods can be applied to the solution of the interface system [3, 4]. Such is not the case, however, for more general operators or irregular domains. For efficiency reasons the system must then be solved by iterative methods, such as the preconditioned conjugate gradient method. Once the solution is known on the interfaces, one more elliptic problem must be solved on each subdomain with the computed values as boundary conditions.

In [3], an eigenvalue decomposition in terms of Fourier modes is given for the capacitance matrix for the case of the Poisson equation on a rectangle divided into two strips. This decomposition is described in section 2. In this paper, we are interested in the analysis of this decomposition, which we will call $M_C$, as a preconditioner on irregular domains and in particular, we want to study the dependency of the convergence rate on the gridsize and the shape of the domain. Many of the preconditioners, when applied to an L-shaped region, have convergence rates that are bounded independently of the gridsize. The bound, however, depends on the relative aspect ratios of the subdomains. For example, all of the preconditioners, except for $M_C$, are known to deteriorate when one of the subdomains becomes narrow. In section 3, we show that, if we use $M_C$ as preconditioner for the capacitance matrix on any L-shaped region, the preconditioned matrix has a condition number that is bounded by 2.16, independently of gridsize and aspect ratios. Given an L-shaped region, there are two ways of separating it into two rectangular subregions. We prove, also in section 3, an interesting property of the preconditioner $M_C$, namely that the convergence rate is not affected by the way we choose to subdivide the domain. In section 4, we discuss the extension of some of the results in section 3 to C-shapes. In the proofs of sections 3 and 4, we often use a common operator, which describes the interaction between two perpendicular interior interfaces. This operator is analyzed in detail in the appendix.

2. The interface operator and its preconditioners

In order to illustrate the method, we will apply the process described above to a simple region $\Omega$, which can be decomposed into two rectangles $\Omega_1$ and $\Omega_2$, with interface $\Gamma_3$, as shown in fig.1. Let the linear system

$$Au = f$$

represent the discretization of the differential operator on $\Omega$. By ordering the variables in $\Omega_1$ and $\Omega_2$ first and then those in $\Gamma_3$, the system (2.1) can be written in block form as:

$$
\begin{pmatrix}
A_{11} & A_{13} \\
A_{21} & A_{22} \\
A_{31} & A_{33}
\end{pmatrix}
\begin{pmatrix}
u_1 \\
u_2 \\
u_3
\end{pmatrix} =
\begin{pmatrix}f_1 \\
f_2 \\
f_3
\end{pmatrix},
$$

(2.2)

where the indexes for $u$ and $f$ corespond to gridpoints in $\Omega_1$, $\Omega_2$ and $\Gamma_3$, respectively. Based on
the following block decomposition of the matrix in (2.2):

$$ A = \begin{pmatrix} A_{11} & A_{12} \\ A_{13} & A_{22} \\ \end{pmatrix} \begin{pmatrix} I & A_{11}^{-1}A_{13} \\ I & A_{22}^{-1}A_{23} \\ \end{pmatrix}, \tag{2.3} $$

where $C$ is the Schur complement of $A_{33}$ in $A$, i.e.

$$ C = A_{33} - A_{13}^T A_{11}^{-1} A_{13} - A_{23}^T A_{22}^{-1} A_{23}, \tag{2.4} $$

the system (2.2) can be solved as follows:

**Step 1:** Compute

$$ g = f_3 - A_{13}^T A_{11}^{-1} f_1 - A_{23}^T A_{22}^{-1} f_2 \tag{2.5} $$

and solve

$$ Cu_3 = g \tag{2.6} $$

**Step 2:** Solve

$$ A_{11} u_1 = f_1 - A_{13} u_3 \tag{2.7} $$

and

$$ A_{22} u_2 = f_2 - A_{23} u_3 \tag{2.8} $$

The computation of $g$ by (2.5) and $u_1$ and $u_2$ by (2.7) and (2.8), require the solution of independent problems on the subdomains. The matrix $C$ given by (2.4), also called the capacitance matrix, is dense and expensive to compute. It is possible, however, to compute the action of $C$ on a vector $v$ at the cost of solving problems on the subdomains with boundary conditions on $\Gamma$ given by $v$. Therefore, the interface system (2.6) is often solved by preconditioned conjugate gradients (PCG). Since each iteration involves solving problems on the subdomains, it is essential to keep the number of iterations low. For this reason, much effort has been devoted recently to the construction of good preconditioners for the capacitance matrix [6, 1, 7, 3, 4]. Many of the preconditioners proposed are spectrally equivalent to the exact boundary operator. They therefore yield convergence rates that are bounded independently of the gridsize. The method is particularly suited to problems for which the subproblems can be solved efficiently, for example, when the operator has separable coefficients. When the subdomain problems cannot be solved efficiently but they can be approximated by separable operators, it is possible to derive block preconditioners for the original system based on preconditioners for the interface system [8, 2, 5].
In [3], the case of a constant coefficient operator on a rectangular domain divided into two strips is analyzed. For this simple case, it is shown that, for many of the preconditioners proposed in the literature, while the condition number of the preconditioned system can be bounded independently of the gridsize $h$ for a fixed domain, it can grow as a function of the aspect ratio of the subdomains. Roughly speaking, the aspect ratio of a rectangle is the ratio between its height and its width (note: for one of the preconditioners proposed in [1], the bound grows when only one of the subdomains becomes narrow). A fast direct solver for $C$ based on Fourier analysis can be derived from the exact eigenvalue decomposition of the capacitance matrix. This operator takes aspect ratios into account and solves exactly the interface problem for the case of constant coefficients on a rectangle divided into two strips. It is therefore proposed in [3] to apply it as a preconditioner for interface systems on irregular regions or for variable coefficient operators. We will call this preconditioner $M_C$. For the case of a five point finite differences discretization of the Poisson equation:

$$-u_{xx} - u_{yy} = f$$  \hspace{1cm} (2.9)

on a regular grid of size $h = \frac{1}{n+1}$, $M_C$ has a decomposition of the form:

$$M_C = W_n \Lambda W_n^T$$  \hspace{1cm} (2.10)

where $\Lambda$ is a diagonal matrix and $W_n$ is the matrix of sine modes of dimension $n$, whose elements are given by:

$$w_{ij} = \sqrt{\frac{2}{n+1}} \sin \frac{ij \pi}{n+1}$$  \hspace{1cm} (2.11)

for $i, j = 1, \ldots, n$.

Given integers $n, m_1$ and $m_2$, define

$$\lambda_j(n, m_1, m_2) = \left( \frac{1 + \gamma_j^{m_1+1}}{1 - \gamma_j^{m_1+1}} + \frac{1 + \gamma_j^{m_2+1}}{1 - \gamma_j^{m_2+1}} \right) \sqrt{\sigma_j + \frac{\sigma_j^2}{4}}$$  \hspace{1cm} (2.12)

where

$$\sigma_j = 4 \sin^2 \left( \frac{j}{(n+1)} \frac{\pi}{2} \right)$$  \hspace{1cm} (2.13)

and

$$\gamma_j = \left( 1 + \frac{\sigma_j}{2} - \sqrt{\sigma_j + \frac{\sigma_j^2}{4}} \right)^2$$  \hspace{1cm} (2.14)

The eigenvalues of $M_C$ are given by

$$\lambda_j = \lambda_j(n, m_1, m_2)$$

for $j = 1, \ldots, n$, where $m_1$ and $m_2$ are the number of grid points in the $y$-direction in $\Omega_1$ and $\Omega_2$ respectively.

The preconditioners proposed in [6] and [7] have the same eigenvectors as (2.10), but the eigenvalues are those of the square root of the one-dimensional discrete Laplacian, namely $\sqrt{\sigma_j}$ in [6] and $\sqrt{\sigma_j + \frac{\sigma_j^2}{4}}$ in [7]. For the case of the Poisson equation (2.9), it can be proved that one of the preconditioners given in [1] also has a decomposition of the form (2.10). The eigenvalues $\lambda_j$ for this operator can be obtained by setting $m_2 = m_1$ in (2.12), i.e. $\lambda_j(n, m_1, m_1)$. This preconditioner
is therefore exact for the case of a rectangle divided symmetrically into two identical rectangular subdomains.

3. L-shaped regions

In this section, we describe the interface operator and its preconditioners for an L-shaped domain, the simplest irregular shape that can be decomposed in rectangular subregions. Consider the Poisson equation on the region Ω of fig.2. It is clear that either interface, Γ₄ or Γ₅, will divide the domain into two rectangles. We might ask ourselves two questions: is a particular decomposition better than the other? And how does the convergence rate depend on the mesh size and the aspect ratios of the subdomains? We will show that for the particular preconditioner \( M_C \) we analyze, the two decompositions produce iteration matrices with the same convergence rate. We also give a bound for the condition number that is independent of the mesh size and the subdomain aspect ratios.

\[ \begin{array}{c}
\text{n} \\
\text{m₁} \\
Ω₁ \\
Γ₄ \\
\text{m₂} \\
Ω₂ \\
Γ₅ \\
Ω₃ \\
\end{array} \]

**Figure 2: L-shaped domain**

Let the linear system

\[ Au = f \]  \hspace{1cm} (3.1)

represent a standard second order five point discretization of the differential equation on a regular grid imposed on the domain Ω. Let us first consider the domain Ω as the union of two rectangles divided by the interface Γ₄. An interface system of the form

\[ C₄u₄ = g₄ \] \hspace{1cm} (3.2)

can be derived for the variables on Γ₄ by the process of block elimination, similarly to equations (2.5) to (2.8).

Similarly, we can consider the domain Ω as the union of two rectangles divided by the interface Γ₅ and an interface system of the form

\[ C₅u₅ = g₅ \] \hspace{1cm} (3.3)

can be derived for the variables on Γ₅.

On the other hand, by reordering the gridpoints on the subdomains first and then those on the interfaces Γ₄ and Γ₅, \( A \) can be written in block form as:

\[ A = \begin{pmatrix} A_Ω & P \\ P^T & A_Γ \end{pmatrix} \] \hspace{1cm} (3.4)
where

\[ A_\Omega = \begin{pmatrix} A_{11} & A_{22} \\ A_{22} & A_{33} \end{pmatrix}, \quad A_\Gamma = \begin{pmatrix} A_{44} & A_{55} \\ A_{55} & A_{66} \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} A_{14} & A_{24} \\ A_{24} & A_{35} \end{pmatrix}. \]

The matrix \( A \) of (3.4) can be decomposed as follows:

\[ A = \begin{pmatrix} A_\Omega & P^T \\ C_{45} & I \end{pmatrix} \begin{pmatrix} I & A^{-1}_\Omega P \\ 0 & I \end{pmatrix}, \quad (3.5) \]

where \( C_{45} \) is the Schur complement of \( A_\Gamma \) in \( A \), i.e.,

\[ C_{45} = A_\Gamma - P^T A^{-1}_\Omega P = \begin{pmatrix} M_4 & -A_{24}^T A_{22}^{-1} A_{24} \\ -A_{25}^T A_{22}^{-1} A_{25} & M_5 \end{pmatrix}, \quad (3.6) \]

with

\[ M_4 = A_{44} - A_{14}^T A_{11}^{-1} A_{14} - A_{24}^T A_{22}^{-1} A_{24} \quad (3.7) \]

and

\[ M_5 = A_{55} - A_{25}^T A_{22}^{-1} A_{25} - A_{35}^T A_{33}^{-1} A_{35}. \quad (3.8) \]

The matrix \( M_4 \) would be the capacitance matrix for \( \Gamma_4 \) if the domain \( \Omega_3 \) were absent. Similarly, \( M_5 \) would be the capacitance matrix for \( \Gamma_5 \) if the domain \( \Omega_1 \) were absent. In fact, they are nothing but the preconditioner \( M_C \) described in the previous section. Both \( M_4 \) and \( M_5 \) have eigenvalue decompositions of the form (2.10). According to the definition (2.12), the eigenvalues of \( M_4 \) are given by \( \lambda_j(n, m_1, m_2) \) for \( j = 1, \ldots, n \) and its eigenvectors, by \( W_n \). The eigenvalues of \( M_5 \) are \( \lambda_i(\ell_2, n, m) \) for \( i = 1, \ldots, m_2 \), with eigenvectors given by \( W_{m_2} \).

The matrix \( C_4 \) of (3.2) is the Schur complement of \( A_{44} \) in \( A \), but it can also be written as the Schur complement of \( M_4 \) in \( C_{45} \). Similarly, \( C_5 \) is the Schur complement of \( A_{55} \) in \( A \), but it can also be written as the Schur complement of \( M_5 \) in \( C_{45} \). Therefore, we can derive the following expressions for \( C_4 \) and \( C_5 \):

**Lemma 3.1.** The interface matrix for \( \Gamma_4 \) in \( \Omega \) can be written as

\[ C_4 = M_4 - B^T M_5^{-1} B, \quad (3.9) \]

where \( B = A_{25}^T A_{22}^{-1} A_{24} \). Similarly, the interface matrix for \( \Gamma_5 \) in \( \Omega \) can be written as

\[ C_5 = M_5 - B M_4^{-1} B^T. \quad (3.10) \]

The preconditioner proposed in [3] for \( C_4 \) in (3.2) would correspond to \( M_C = M_4 \) and similarly, \( M_C = M_5 \) for \( C_5 \) in (3.3). Since \( M_C \) is positive definite, we can choose \( \sqrt{M_C} \) as a symmetric preconditioner. Let us define the preconditioned matrices:

\[ \hat{C}_4 = M_4^{-1/2} C_4 M_4^{-1/2} \quad \text{and} \quad \hat{C}_5 = M_5^{-1/2} C_5 M_5^{-1/2}. \quad (3.11) \]

By (3.9), we have

\[ \hat{C}_4 = I_n - \hat{B}^T \hat{B} \quad \text{and} \quad \hat{C}_5 = I_{m_2} - \hat{B} \hat{B}^T. \quad (3.12) \]
where

$$
\hat{B} = M_5^{-1/2} A_{52}^T A_{22}^{-1} A_{24} M_4^{-1/2} .
$$  \hspace{1cm} (3.13)

If we choose \( \Gamma_4 \) as the interface, at each iteration subdomain problems will be solved on \( \Omega_1 \) and \( \Omega_2 \cup \Gamma_5 \cup \Omega_3 \). Similarly, if we choose \( \Gamma_5 \) as the interface, at each iteration subdomain problems will be solved on \( \Omega_1 \cup \Gamma_4 \cup \Omega_2 \) and \( \Omega_3 \). The work per iteration is therefore comparable for both ways of splitting the domain. We will next show that, by solving (3.2) with preconditioner \( M_4 \) and (3.3) with preconditioner \( M_5 \), both systems are also equivalent from the convergence point of view. Therefore, in a general case, there is no \textit{a priori} reason to prefer one way of decomposing the domain over the other.

If \( n = m_2 \), this fact is not surprising, considering that both interface systems have the same order and it is easy to see that \( \hat{C}_4 = \hat{C}_5 \). It is not obvious, however, whether one way of decomposing the domain should be preferred when \( n \neq m_2 \). As it turns out, even in this case the asymptotic convergence rate is the same for both systems, because the matrix \( \hat{C}_{45} \) of (3.6) satisfies the following theorem:

\textbf{Theorem 3.1.} Consider the following symmetric positive definite (SPD) system, written in block form:

$$
\begin{pmatrix}
A & B \\
B^T & C
\end{pmatrix}
\begin{pmatrix}
x \\
y
\end{pmatrix} =
\begin{pmatrix}
f \\
g
\end{pmatrix},
$$  \hspace{1cm} (3.14)

where the blocks \( A \) and \( B \) are square matrices. Also, define the Schur complement systems:

$$
(A - BC^{-1}B^T)x = f - BC^{-1}g
$$

and

$$
(C - B^T A^{-1}B)y = g - B^T A^{-1}f .
$$  \hspace{1cm} (3.15)

Consider the solution of (3.15) by the following fixed-point iteration, with splitting matrix given by \( A \): given an initial guess \( x^0 \), define the \( i \)-th iterate as the solution to:

$$
Ax^i = f - BC^{-1}g + BC^{-1} B^T x^{i-1}
$$

for \( i = 1, 2, \ldots \)

Similarly, given \( y^0 \), define the \( i \)-th iterate of a fixed-point iteration for solving (3.16) with splitting matrix given by \( C \), as:

$$
Cy^i = g - B^T A^{-1} f + B^T A^{-1} B y^{i-1}
$$

for \( i = 1, 2, \ldots \)

Then, the two iterations are convergent. Moreover, they are equivalent in the sense that for any given initial guess \( x^0 \) for (3.17), there exists an initial guess \( y^0 \) for (3.18), such that for all \( i = 0, 1, \ldots \) we have:

$$
y^i = q_y + P_y x^i ,
$$

where \( q_y = C^{-1}g \) and \( P_y = -C^{-1} B^T \) and

$$
\|e_x^{i+1}\|_A \leq \|e_x^i\|_A \leq \|e_x^0\|_A ,
$$

where \( e_x^i = x^i - x , e_y^i = y^i - y \) and \( \|u\|_A \) denotes the \( A \)-norm of a vector \( u \), i.e. \( \sqrt{u^T A u} \).

Completely analogous results also hold for \( x^0 \) and \( e_x^0 \), given an initial guess \( y^0 \) for (3.16).

\textbf{Proof.} Given \( x^0 \), define \( y^0 = q_y + P_y x^0 \). By induction, we can see that (3.19) satisfies (3.18) for every \( i \geq 1 \).
From the classical matrix iterative analysis for the convergence of block Gauss-Seidel iteration for SPD matrices, it can be shown that the two iterations converge. Also, since the matrix of (3.14) is SPD, so are the blocks \( A \) and \( C \) and their corresponding Schur complements. Therefore, \( A^{-1/2} \) and \( C^{-1/2} \) are well defined and
\[
\|A^{-1/2}BC^{-1/2}\|_2 \leq 1 .
\] (3.21)
We can also prove that
\[
Ae_x^{i+1} = BC^{-1}B^T e_x^i
\]
and
\[
e_y^i = P_y e_x^i .
\]
Therefore, we have
\[
A^{1/2}e_x^{i+1} = -(A^{-1/2}BC^{-1/2})C^{1/2}e_y^i
\] (3.22)
and
\[
C^{1/2}e_y^i = -(C^{-1/2}B^T A^{-1/2})A^{1/2}e_x^i
\] (3.23)
Using (3.21), (3.22) and (3.23), we can prove (3.20).

When an iterative method such as PCG is used, the rate of convergence depends on the condition number of the corresponding preconditioned matrix in (3.11). By applying the last theorem to \( C_{45} \) and by using (3.12), we can conclude the following:

**Theorem 3.2.** Solving both systems (3.2) and (3.3) with preconditioners of the form \( M_C \) produce equivalent asymptotic convergence rates. Moreover, by (3.12), we have
\[
\kappa(\tilde{C}_4) \leq \frac{1}{1 - \|\tilde{B}^T \tilde{B}\|_2}
\] (3.24)
and
\[
\kappa(\tilde{C}_5) \leq \frac{1}{1 - \|\tilde{B}^T \tilde{B}\|_2} .
\] (3.25)

Numerical computations show that the singular values \( \beta_i \) of \( \tilde{B} \) decrease very quickly with the index \( i \). Therefore, in practice, only a few eigenvalues of \( \tilde{C}_4 \) and \( \tilde{C}_5 \) are different from 1, which leads to rapid convergence of the PCG method when applied to either matrix. For example, for the L-shaped region with corners at: \((0, 0), (3, 0), (3, 0.25), (1, 0.25), (1, 1.25)\) and \((0, 1.25)\), for \( n = 31 \) and \( 63 \), table 1 shows the singular values of \( \tilde{B} \) and the eigenvalues of \( \tilde{C}_5 \), computed in single precision.

Our conclusion is that either way of decomposing an L-shaped region into two rectangles produces the same convergence rate, when preconditioner \( M_C \) is used. Moreover, we will be able to give an analytical bound on the condition number of the preconditioned capacitance matrix. This bound is derived from a bound on the norm of the operator \( \tilde{B}^T \tilde{B} \).

But first, we will give an expression for the elements of a unitary transformation of \( \tilde{B} \). Let the elements of the matrix \( W_n \) be given by (2.11) and similarly, define the elements of \( W_{m_2} \) by replacing \( n \) by \( m_2 \) in (2.11).

The operator
\[
Q_{NE} = A_{22}^T A_{22}^{-1} A_{24} ,
\]
which is part of the definition (3.13) of \( \tilde{B} \), is the operator that takes boundary values on the interface \( \Gamma_4 \), solving a Poisson problem on \( \Omega_2 \) and then takes the values of the solution at the
gridpoints which are adjacent to \( \Gamma_5 \). It is possible to derive the elements of \( Q_{NE} \) when it is pre and post-multiplied, respectively, by the matrices \( W_{m_2} \) and \( W_n \). The elements of \( W_{m_2}Q_{NE}W_n \) are given by

\[
q_{ij} = \frac{2}{\sqrt{(m_2+1)(n+1)}} \frac{\sin \frac{i\pi}{m_2+1} \sin \frac{j\pi}{n+1}}{s_j^{(m_2)} \sigma_i^{(n)} + s_i^{(m_2)}} \tag{3.26}
\]

for \( i = 1, \ldots, m_2 \) and \( j = 1, \ldots, n \). A proof of (3.26) can be found in the appendix (see lemma 5.2).

For any given integers \( n, m_1 \) and \( m_2 \), let \( \lambda(n, m_1, m_2) \) be defined by (2.12), where \( \gamma_j \) is given by equation (2.14). By using (3.26), it is easy to prove the following lemma:

**Lemma 3.2.** Let

\[
V = W_{m_2} \hat{B} W_n \tag{3.27}
\]

Then, \( \|V\|_2 = \|\hat{B}\|_2 \) and the elements of the matrix \( V \) are given by

\[
v_{ij} = \frac{2}{\sqrt{(n+1)(m_2+1)}} \frac{\sin \frac{i\pi}{m_2+1} \sin \frac{j\pi}{n+1}}{s_j^{(m_2)} \sigma_i^{(n)} + s_i^{(m_2)}} \tag{3.28}
\]

for \( i = 1, \ldots, m_2 \) and \( j = 1, \ldots, n \), where \( s_j^{(4)} = \sqrt{\lambda_j(n, m_1, m_2)} \) and \( s_i^{(5)} = \sqrt{\lambda_i(m_2, n, m_3)} \).

As equations (3.24) and (3.25) suggest, in order to find a bound for the condition number of the preconditioned capacitance system, we need to bound the norm of \( \hat{B} \), or \( V \). Since we have an expression for the elements of \( V \), we can bound \( \|V\|_1 \) and \( \|V\|_\infty \) and then use the property:

\[\|V\|_2 \leq \sqrt{\|V\|_1 \|V\|_\infty} \]

The results are summarized in the next theorem. A proof can be found in appendix B:
Theorem 3.3. Define the aspect ratio for the domain $\Omega_2$ in fig.2 as $\alpha = \frac{m_2 + 1}{n_1}$. Then,

\begin{align*}
\text{a) } &\|V\|_1 \leq \sqrt{\alpha} 0.733 \text{ and } \|V\|_\infty \leq \sqrt{\frac{1}{\alpha}} 0.733. \\
\text{b) } &\|\bar{B}^T \bar{B}\|_2 \leq \|B\|_2^2 = \|V\|_2^2 \leq \|V\|_1 \|V\|_\infty \leq 0.54. \\
\text{c) } &\text{For all gridsizes and all L-shaped regions,} \\
&\mathcal{K}(\bar{C}_4) \leq 2.16 \text{ and } \mathcal{K}(\bar{C}_6) \leq 2.16. \\
\end{align*} (3.29)

In our experiments on L-shaped domains with many different aspect ratios, condition numbers larger than 1.2 have not been observed. The bound 0.54 in (b), however, is fairly tight for $\|V\|_1 \|V\|_\infty$, as was shown by numerical experiments with large values of $n$ and $m_2$. Therefore, if a tighter bound is desired for the condition number, one would need to bound the 2-norm of $\bar{B}^T \bar{B}$ directly.

We would also like to discuss briefly how the parameter $n_3$ (or, respectively, $m_1$) affects the performance of preconditioner $M_4$ ($M_6$). Clearly, as $n_3$ tends to zero for large $m_2$, the domain $\Omega$ approaches the shape of a perfect rectangle. The preconditioner $M_4$ should reflect this by becoming the exact boundary operator. In other words, $\mathcal{K}(\bar{C}_4)$ should approach one. We can verify that this is the case as follows: $\nu_{ij}$ in (3.28) depends on $n_3$ only through $\lambda_i(m_2, n, n_3)$ (defined in (2.12)). When the aspect ratio $\frac{n_3 + 1}{m_2 + 1}$ tends to zero (i.e. $\Omega_3$ becomes thinner), $\lambda_i(m_2, n, n_3)$ tends to infinity and therefore $\nu_{ij}$ tends to zero. However, we can see that this dependency is very weak, because $\lambda_j(m_2, n, n_3)$ tends rapidly to an asymptotic value independent of $n_3$ when such aspect ratio grows. Only the fact that

$$
\lambda_j(m_2, n, n_3) \geq 2\sqrt{\sigma_j}
$$

is used in the proof of theorem 3.3, which is true for all $n_3$. The discussion above implies that the performance of $M_4$ as a preconditioner for $C_4$ is fairly independent on how irregular the region is.

Incidentally, for the other preconditioners mentioned in this paper [6, 1, 7], the preconditioned capacitance matrix always has the form $X + \bar{B}^T \bar{B}$, for some operator $\bar{B}$ to which the bounds (a) and (b) of theorem 3.3 can also be applied, as long as (3.30) holds. The bound given in (c), however, does not hold for other preconditioners, for which the norm of $X$ may grow when the aspect ratio $\alpha$ of the domain $\Omega_2$ decreases (see [3] for an example on a T-shaped region).

4. C-shaped regions

Some of the expressions and results of the previous section are more general than they appear and they can be used as basic components for more complicated regions that are unions of rectangles. For example, a C-shaped region can be subdivided as indicated in fig.3.

Similar to L-shaped domains, the region of fig.3 can be separated in three rectangles by either $\Gamma_6$ and $\Gamma_7$, or $\Gamma_8$ and $\Gamma_9$. By ordering the variables in $\Omega_i, i \leq 5$ first and then those on $\Gamma_j, 6 \leq j \leq 9$, the matrix $A$ that represents the discrete differential operator on $\Omega$ can be written in block form as in (3.4), where

$$
A_\Omega = \begin{pmatrix} A_{11} & \cdots & \cdots \\ \cdots & A_{55} & \cdots \\ \cdots & \cdots & \cdots \\ \end{pmatrix}, \quad A_\Gamma = \begin{pmatrix} A_{66} & \cdots \\ \cdots & \cdots \\ \cdots & \cdots & A_{99} \end{pmatrix},
$$

(4.1)

and

$$
P = \begin{pmatrix} A_{16} & A_{18} \\ A_{26} & A_{27} & \cdots \\ \cdots & \cdots & \cdots \\ & & & A_{59} \end{pmatrix}.
$$
A system
\[
C_{67} \begin{pmatrix} u_6 \\ u_7 \end{pmatrix} = g_{67}
\] (4.2)
can be derived by block elimination for the interfaces $\Gamma_6$ and $\Gamma_7$, where $C_{67}$ is the Schur complement in $A$ of the blocks $A_{66}$ and $A_{77}$. In [4], a multistrip operator $M_{67}$ is described, which solves, exactly, the problem on a rectangle divided into three strips ($\Omega_1$, $\Omega_2$ and $\Omega_3$). We will analyze $M_{67}$ as a preconditioner for $C_{67}$. The operator $M_{67}$ has the following block structure:
\[
M_{67} = \begin{pmatrix} H_6 & S \\ S & H_7 \end{pmatrix},
\] (4.3)
where
\[
H_6 = A_{66} - A_{16}^T A_{11}^{-1} A_{16} - A_{26}^T A_{22}^{-1} A_{26},
\]
\[
H_7 = A_{77} - A_{37}^T A_{33}^{-1} A_{37} - A_{27}^T A_{22}^{-1} A_{27}
\]
and
\[
S = -A_{26}^T A_{22}^{-1} A_{27}.
\]
The blocks $H_6$, $H_7$ and $S$ have eigenvalue decompositions of the form (2.10), with eigenvalues given by $\lambda_j(n, m_1, m_2)$, $\lambda_j(n, m_2, m_3)$ and
\[
\delta_j(n, m_2) = -2 \sqrt{\sigma_j + \frac{\sigma_j^2}{4} \left( \frac{\gamma_j^{m_2+1}}{1 - \gamma_j^{m_2+1}} \right)}
\] (4.4)
respectively, for $j = 1, \ldots, n$. (See lemma 5.1 in appendix A).

Similarly, a system
\[
C_{89} \begin{pmatrix} u_8 \\ u_9 \end{pmatrix} = g_{89}
\] (4.5)
can be derived for the interfaces $\Gamma_8$ and $\Gamma_9$, where $C_{89}$ is the Schur complement in $A$ of the blocks $A_{88}$ and $A_{99}$. The system (4.5) can be preconditioned by a block diagonal preconditioner $M_{89}$, with diagonal blocks $M_6$ and $M_9$. $M_8$ is the exact interface system for $\Gamma_8$ with respect to the subdomains
\( \Omega_1 \) and \( \Omega_4 \), and \( M_9 \) is the exact interface system for \( \Gamma_9 \) with respect to the subdomains \( \Omega_3 \) and \( \Omega_5 \). Both \( M_8 \) and \( M_9 \) have decompositions of the form (2.10).

It can be easily shown that \( C_\Gamma \), the Schur complement of the blocks \( A_\Gamma \) in \( A \), can be written

\[
C_\Gamma = \begin{bmatrix}
M_{67} & Q_{SE} & 0 \\
Q_{SE}^T & 0 & Q_{NE} \\
0 & Q_{NE}^T & M_8 \\
\end{bmatrix},
\]

where \( Q_{SE} = A_{16}^T A_{11}^{-1} A_{18} \) and \( Q_{NE} = A_{37}^T A_{33}^{-1} A_{39} \). Again, by applying theorem 3.1, we have that both ways of dividing the domain are equivalent, in the sense that initial residuals can be found such that the same number of iterations are necessary when PCG is applied to \( C_{67} \) with preconditioner \( M_{67} \) than when PCG is applied to \( C_{89} \) with preconditioner \( M_{89} \). The preconditioned interface operator for \( \Gamma_6 \) and \( \Gamma_7 \),

\[
\hat{C}_{67} \equiv M_{67}^{-1/2} C_{67} M_{67}^{-1/2},
\]

can be written in the form

\[
\hat{C}_{67} = I - \hat{B}^T \hat{B},
\]

where \( \hat{B} \in \mathbb{R}^{(m_1+m_3) \times 2n} \) and

\[
\hat{B} = \begin{pmatrix}
M_8 & 0 \\
0 & M_9
\end{pmatrix}^{-1/2} \begin{pmatrix}
Q_{SE}^T & 0 \\
0 & Q_{NE}^T
\end{pmatrix} M_{67}^{-1/2}.
\]

Similarly, the preconditioned interface operator for \( \Gamma_8 \) and \( \Gamma_9 \),

\[
\hat{C}_{89} \equiv \begin{pmatrix}
M_8 & 0 \\
0 & M_9
\end{pmatrix}^{-1/2} C_{89} \begin{pmatrix}
M_8 & 0 \\
0 & M_9
\end{pmatrix}^{-1/2},
\]

can be written in the form

\[
\hat{C}_{89} = I - \hat{B} \hat{B}^T.
\]

The condition numbers of \( \hat{C}_{67} \) and \( \hat{C}_{89} \) are bounded by

\[
\kappa(\hat{C}_{67}) \leq \frac{1}{1 - \|\hat{B}^T \hat{B}\|_2} \quad (4.8)
\]

and

\[
\kappa(\hat{C}_{89}) \leq \frac{1}{1 - \|\hat{B}^T \hat{B}\|_2} \quad (4.9)
\]

Define \( V \) as the following unitary transformation of \( \hat{B} \):

\[
V = \begin{pmatrix}
W_{m_1} & 0 \\
0 & W_{m_3}
\end{pmatrix} \hat{B} \begin{pmatrix}
W_n & 0 \\
0 & W_n
\end{pmatrix}.
\]

Then \( \|V\| = \|\hat{B}\| \). The matrix \( V \) can be written as a block two by two matrix

\[
V = \begin{pmatrix}
V_{66} & V_{67} \\
V_{76} & V_{77}
\end{pmatrix},
\]

(4.10)
whose block elements have expressions similar to the matrix \( V \) for L-shaped regions, namely,

\[
\begin{align*}
V_{66} &= \frac{1}{2} W_{m_1} M_8^{-1/2} Q_{SE}^T W_n R_6 \\
V_{67} &= \frac{1}{2} W_{m_1} M_8^{-1/2} Q_{SE}^T W_n R_- \\
V_{76} &= \frac{1}{2} W_{m_3} M_9^{-1/2} Q_{NE}^T W_n R_- \\
V_{77} &= \frac{1}{2} W_{m_3} M_9^{-1/2} Q_{NE}^T W_n R_7
\end{align*}
\tag{4.11}
\]

where \( R_6, R_7 \) and \( R_- \) are diagonal matrices such that:

\[
\begin{pmatrix} R_6 & R_- \\ R_- & R_7 \end{pmatrix} = \begin{pmatrix} W_n & 0 \\ 0 & W_n \end{pmatrix} M_8^{-1/2} \begin{pmatrix} W_n & 0 \\ 0 & W_n \end{pmatrix}
\]

For the case when \( m_1 = m_3 \leq m_2 \), a simple expression can be found for \( R_6, R_7 \) and \( R_- \), namely \( R_6 = R_7 = R_+ \), with the diagonal elements of \( R_\pm \) given by

\[
r_\pm^j = \frac{1}{2} \left( \frac{1}{\sqrt{\lambda_j} - |\delta_j|} \pm \frac{1}{\sqrt{\lambda_j} + |\delta_j|} \right),
\]

where \( \lambda_j \) is \( \lambda_j(n, m_1, m_2) \), given by (2.12) and \( \delta_j \) is \( \delta_j(n, m_2) \), given by (4.4). Arguments similar to those in theorem 3.3 can be applied to give the following:

**Theorem 4.1.** Consider a C-shaped region like fig. 3, where \( m_1 = m_3 \leq m_2 \) and \( \alpha \) is the aspect ratio for the domain \( \Omega_1 \) or \( \Omega_3 \) in the picture, i.e. \( \alpha = \frac{m_1 + 1}{n+1} \). Then,

a) \( \| V \|_1 \leq \sqrt{\alpha} \times 0.7877 \) and \( \| V \|_\infty \leq \sqrt{\alpha} \times 0.7877 \).

b) \( \| VTV \|_2 \leq \| V \|_2^2 = \| B \|_2^2 \leq \| B \|_1 \| B \|_\infty \leq 0.62 \).

c) \( \mathcal{K}(\hat{C}_{67}) \leq 2.63 \) and \( \mathcal{K}(\hat{C}_{89}) \leq 2.63 \) for all gridsizes and all C-shaped regions such that \( m_1 = m_3 \leq m_2 \).

**Proof.** In appendix B.

5. **Appendix A: The interaction between interior edges**

In this appendix, we define the operators that represent the interaction between two interfaces of a given subdomain. Consider the rectangular region \( R \) of fig. 4, with edges \( \Gamma_N, \Gamma_E, \Gamma_S \) and \( \Gamma_W \). This region \( R \) represents a generic rectangular subdomain in the domain \( \Omega \). Let \( n_1 \) be the number of gridpoints in \( \Gamma_N \) (or \( \Gamma_S \)) and \( n_2 \), the number of gridpoints in \( \Gamma_E \) (or \( \Gamma_W \)). The corner points are not included in the edges. They may or may not be interior to \( \Omega \).
For the case of constant coefficients operators, it is possible to describe, in terms of Fourier modes, an operator $Q$ which takes boundary values on one of the edges and computes the solution on the gridpoints adjacent to the same or other edge.

Let $A$ be the matrix which represents the discretization of the differential operator in $\Omega$. If the interface $\Gamma_k$, where $k = N, S, E$ or $W$, is interior to $\Omega$, then we can define $P_k$, the submatrix of $A$ that represents the coupling between gridpoints of $R$ and gridpoints on $\Gamma_k$. Also define $A_R$, the diagonal block corresponding to the interior points of $R$, in other words, $A_R$ is the restriction of the differential operator to the region $R$. We can now define the operator $Q_{kl}$ which represents the interaction between the edges $\Gamma_k$ and $\Gamma_l$ as:

$$Q_{kl} = P_k^T A_R^{-1} P_l$$

(5.1)

For constant coefficients operators, when $\Gamma_k$ and $\Gamma_l$ are parallel, the operator $Q_{kl}$ is diagonalizable by Fourier modes. For example, for the case of the Poisson equation we can prove the following lemma. Here, for any given $n$, $W_n$ is the matrix of sine modes of dimension $n$, with elements given by

$$w_{ij} = \sqrt{\frac{2}{n+1}} \sin \frac{ij\pi}{n+1}$$

(5.2)

for $i, j = 1, \ldots, n$.

**Lemma 5.1.** Consider the Poisson equation on a domain $\Omega$ which contains the rectangular region $R$. Let $Q_{NS}$ be the operator that represents the coupling between interfaces $\Gamma_N$ and $\Gamma_S$, defined as in (5.1). Then,

$$W_n Q_{NS} W_n = D_{NS}$$

where the matrix $D_{NS}$ is diagonal, with diagonal entries given by

$$d_{jj} = \sqrt{\frac{\pi^2}{\gamma_j^2} \frac{1 - \gamma_j}{1 - \gamma_j^{n_2+1}}}$$

(5.3)

where

$$\gamma_j = \left(1 + \frac{\sigma_j^{(1)}}{2} - \sqrt{\sigma_j^{(1)} + \frac{\left(\sigma_j^{(1)}\right)^2}{4}}\right)^2$$

(5.4)

and

$$\sigma_j^{(1)} = 4 \sin^2 \frac{j\pi}{2(n_1 + 1)}$$

(5.5)

A similar expression can be found for $Q_{EW}$.

Also,

$$W_n Q_{NN} W_n = D_{NN}$$

where the matrix $D_{NN}$ is diagonal, with diagonal entries given by

$$d_{jj} = -\sqrt{\gamma_j^2 \frac{1 - \gamma_j^{n_2}}{1 - \gamma_j^{n_2+1}}}$$

(5.6)

Similar expressions can be derived for $Q_{SS}, Q_{EE}$ and $Q_{WW}$.

**Proof.** Proofs for formulas (5.3) and (5.6) can be found in [3] and [4]. Here we give a different – more general – proof using direct (or tensor) products. The matrices $P_N$ and $P_S$ can be written as:

$$P_N = e_1^{(2)} \otimes I_1$$

(5.7)
\[ P_S = e_{n_2}^{(2)} \otimes I_1 \]  \hspace{1cm} (5.8)

where \( I_i \), for \( i = 1, 2 \), is the identity matrix of dimension \( n_i \) and \( e_{i}^{(l)} \) is the \( i \)-th column of \( I_i \). The matrix \( A_R \) is the discrete Laplacian operator on the region \( R \) and it has the following block tridiagonal form:

\[ A_R = \begin{pmatrix} T & I_1 & & \\ I_1 & T & \ddots & \\ & \ddots & \ddots & I_1 \\ I_1 & & I_1 & T \end{pmatrix} \]  \hspace{1cm} (5.9)

where \( T = \text{tridiag}(1, -4, 1) \). It is easy to prove that

\[ W_{n_1}T W_{n_1} = D_T \]

where \( D_T = \text{diag}(-2 - \sigma_j^{(1)}) \). Then we have

\[ W_{n_1}Q_{NS}W_{n_1} = (e_{1}^{(2)} \otimes I_1) \begin{pmatrix} D_T & I_1 & & \\ I_1 & D_T & \ddots & \\ & \ddots & \ddots & I_1 \\ I_1 & & I_1 & D_T \end{pmatrix}^{-1} (e_{n_2}^{(2)} \otimes I_1) \]  \hspace{1cm} (5.10)

By reordering the equations in (5.10), we have:

\[ W_{n_1}Q_{NS}W_{n_1} = (I_1 \otimes e_{1}^{(2)})^T \begin{pmatrix} T_1 & & \\ & T_2 & \ddots \\ & & \ddots \\ & & & T_{n_1} \end{pmatrix}^{-1} (I_1 \otimes e_{n_2}^{(2)}) \]

where \( T_j = \text{tridiag}(1, -2 - \sigma_j^{(1)}, 1) \). Therefore, \( W_{n_1}Q_{NS}W_{n_1} \) is diagonal and its diagonal elements are given by

\[ e_{1}^{(2)^T} T_j^{-1} e_{n_2}^{(2)} \]

which can be proved to be given by (5.3).

Similarly, we can prove that \( W_{n_1}Q_{NN}W_{n_1} \) is diagonal and its diagonal elements are given by

\[ e_{n_2}^{(2)^T} T_j^{-1} e_{n_2}^{(2)} \]

which can be proved to be given by (5.6).

Operators like \( Q_{NE} \), on the other hand, which represent the interaction between perpendicular edges, are not diagonalizable by Fourier modes. Moreover, they are, in general not square, but \( n_1 \) by \( n_2 \) rectangular matrices. It is possible, however, to describe the elements of the matrices \( W_{n_1}Q_{NE}W_{n_2} \) and \( W_{n_1}Q_{NW}W_{n_2} \) for constant coefficients cases.

**Lemma 5.2.** Consider the Poisson equation on a domain \( \Omega \) which contains the rectangular region \( R \). The elements of

\[ \hat{Q}_{NE} = W_{n_1}Q_{NE}W_{n_2} \]

14
are given by
\[ q_{ij}^{NE} = \frac{2}{\sqrt{(n_1 + 1)(n_2 + 1)}} \frac{\sin \frac{j\pi}{n_1 + 1}}{\sigma_i^{(1)} + \sigma_j^{(2)}} \] \hfill (5.11)

for \( i = 1, \ldots, n_1 \) and \( j = 1, \ldots, n_2 \), where \( \sigma_j^{(l)} = 4\sin^2 \frac{j\pi}{2(n_l + 1)} \), for \( l = 1, 2 \). Similarly, the elements of the matrix

\[ \hat{Q}_{NW} = W_{n_1} Q_{NW} W_{n_2} \]

are given by
\[ q_{ij}^{NW} = \frac{2}{\sqrt{(n_1 + 1)(n_2 + 1)}} \frac{\sin \frac{j\pi}{n_1 + 1}}{\sigma_i^{(1)} + \sigma_j^{(2)}} \] \hfill (5.12)

Proof. The eigenvalue decomposition of the matrix \( A_R \) is well known and it is given by
\[ A_R = (W_{n_2} \otimes W_{n_1}) A (W_{n_2} \otimes W_{n_1}) \] \hfill (5.13)

where \( A \) is the \( n_1 n_2 \times n_1 n_2 \) diagonal matrix whose diagonal elements are
\[ \lambda_j = -\sigma_i^{(1)} - \sigma_j^{(2)} \]

with \( J = (j - 1)n_1 + i \), for \( i = 1, \ldots, n_1 \) and \( j = 1, \ldots, n_2 \). Also, we have
\[ P_W = I_2 \otimes e_1^{(1)} \] \hfill (5.14)

By replacing equations (5.13) to (5.14) in (5.1) and then applying the following two properties of tensor products:

i) \( (X \otimes Y)^T = X^T \otimes Y^T \) and

ii) \( (X_1 \otimes Y_1)(X_2 \otimes Y_2) = (X_1 X_2) \otimes (Y_1 Y_2) \)

we have:
\[ Q_{NW} = \left( (e_1^{(2)T} W_2) \otimes W_1 \right) A^{-1} \left( W_2 \otimes (W_1 e_1^{(1)}) \right) \] \hfill (5.15)

and therefore,
\[ \hat{Q}_{NW} = \left( (e_1^{(2)T} W_2) \otimes I_1 \right) A^{-1} \left( I_2 \otimes (W_1 e_1^{(1)}) \right) \] \hfill (5.16)

Then we can see that the \( j \)-th column of (5.16) is given by
\[ \sqrt{\frac{2}{n_2 + 1}} \sin \frac{j\pi}{n_2 + 1} \left( \sigma_j^{(2)} I_1 + \text{diag}(\sigma_i^{(1)}) \right)^{-1} W_1 e_1^{(1)} \]

from which (5.11) follows.

Similarly, (5.12) can be derived by using
\[ P_E = I_2 \otimes e_1^{(1)} \] \hfill (5.17)

instead of (5.7).

6. Appendix B: Proof of theorems 3.3 and 4.1

Proof of Theorem 3.3
Proof.

Theorem 3.3: Let the aspect ratio for the domain $\Omega_2$ in fig.2 be defined as $\alpha = \frac{m_2+1}{n_1}$. Then,

a) $\|\tilde{V}\|_1 \leq \sqrt{\alpha} \cdot 0.733$ and $\|\tilde{V}\|_\infty \leq \sqrt{\frac{1}{\alpha}} \cdot 0.733$.

b) $\|V^T V\|_2 \leq \|\tilde{V}\|_2^2 \leq \|\tilde{V}\|_1 \|\tilde{V}\|_\infty \leq 0.54$.

c) For all grid sizes and all $L$-shaped regions,

\[ \mathcal{K}(\tilde{\mathcal{C}}_4) \leq 2.16 \quad \text{and} \quad \mathcal{K}(\tilde{\mathcal{C}}_5) \leq 2.16 \quad . \] (6.1)

Proof. (b) follows from (a). (c) follows from (b) and from equations (3.24) and (3.25). In order to prove (a), we will first need to prove some lemmas that give bounds for the column and row sums of the absolute values of (3.28), the elements $u_{ij}$ of the matrix $\tilde{V}$. The eigenvalues (2.12) of $M_4$ and $M_5$ can be bounded by

\[ \lambda_j(n, m_1, m_2) \geq 4 \sin \frac{j\pi}{2(n+1)} \] (6.2)

for all $j = 1, \ldots, n$ and

\[ \lambda_i(m_2, n, n_3) \geq 4 \sin \frac{i\pi}{2(m_2+1)} \] (6.3)

for all $i = 1, \ldots, m_2$. It is easy to show that

\[ |v_{ij}| \leq \frac{1}{2\sqrt{\alpha} \cdot \frac{n+1}{n+1}} \] (6.4)

where the function $f$ is defined by

\[ f(x, y) = \frac{\sqrt{\sin x \frac{\pi}{2}} \cdot \cos x \frac{\pi}{2} \cdot \sqrt{\sin y \frac{\pi}{2}} \cdot \cos y \frac{\pi}{2}}{\sin^2 x \frac{\pi}{2} + \sin^2 y \frac{\pi}{2}} \] (6.5)

$x_i = \frac{i}{m_2+1}$ and $y_j = \frac{j}{n+1}$. Similarly, we have

\[ |v_{ij}| \leq \frac{\sqrt{\alpha}}{2 \cdot \frac{m_2+1}{m_2+1}} \] (6.6)

The column and row sums of $|v_{ij}|$ can be then bounded by expressions that involve the integrals of $f$ with respect to $x$ or with respect to $y$. The following lemma gives an expression for the integral of $f$ with respect to $x$, for a fixed $y$. Since $f(x, y) = f(y, x)$, an analogous result holds for the integral of $f$ with respect to $y$, for a fixed $x$.

Lemma 6.1. Given $y \in (0, 1)$ and $a, b \in (0, 1)$ such that $a \leq y < b$,

\[ \int_a^b f(x, y) \, dx = \frac{2}{\sqrt{2\pi}} \cos y \frac{\pi}{2} \left( \pi - g(\sin b \frac{\pi}{2}, \sin y \frac{\pi}{2}) + g(\sin a \frac{\pi}{2}, \sin y \frac{\pi}{2}) \right) \] (6.7)

where

\[ g(z, w) = \frac{1}{2} \log \frac{z + \sqrt{2zw} + w}{z - \sqrt{2zw} + w} + \arctan \frac{\sqrt{2zw}}{z - w} \] (6.8)
Proof. By replacing \( z = \sin x \frac{\pi}{2} \) and \( w = \sin y \frac{\pi}{2} \) in (6.5) and defining
\[
F(z, w) = \frac{\sqrt{zw}}{z^2 + w^2} ,
\]
we get
\[
\int_a^b f(x, y) \, dx = \frac{2}{\pi} \cos y \frac{\pi}{2} \int_a^{\sin \frac{y}{2}} F(z, w) \, dz
\]
\[
= \frac{2}{\sqrt{2\pi}} \cos y \frac{\pi}{2} \left( -\frac{1}{2} \log \frac{z + \sqrt{2zw + w}}{z - \sqrt{2zw + w}} + \arctan \frac{\sqrt{2zw}}{w - z} \right) \bigg|_{\sin \frac{b}{2}}^{\sin \frac{a}{2}}
\]
\[
= \frac{2}{\sqrt{2\pi}} \cos y \frac{\pi}{2} \left( \pi - g(\sin b \frac{\pi}{2}, \sin y \frac{\pi}{2}) + g(\sin a \frac{\pi}{2}, \sin y \frac{\pi}{2}) \right) .
\]

We will also need to describe the behavior of \( f(x, y) \), for a fixed \( y \in (0, 1) \), in the interval \( x \in (0, 1) \). We can easily see that, when \( x, y \in (0, 1) \), \( f(x, y) > 0 \). In the next lemma we prove that \( f(\cdot, y) \) has one and only one relative maximum in \((0, 1)\).

Lemma 6.2. Given \( y \in (0, 1) \), there exists a unique \( x^*(y) \in (0, 1) \) such that:
\[
\max_{0 < \xi < 1} f(x, y) = f(x^*, y) ,
\]
\( f(\cdot, y) \) is monotonically increasing on the interval \((0, x^*)\) and \( f(\cdot, y) \) is monotonically decreasing on \((x^*, 1)\). Moreover, \( f(x^*, y) \) is bounded by
\[
f(x^*, y) \leq \frac{3}{4\sqrt{3}} \cot y \frac{\pi}{2} .
\]

Proof. The partial derivative of \( f \) with respect to \( x \) is given by:
\[
\frac{\partial f}{\partial x} = \xi(x, y)(\sin^2 x \frac{\pi}{2} - z_-)(\sin^2 x \frac{\pi}{2} - z_+) ,
\]
where \( \xi(x, y) > 0 \) for all \( x, y \in (0, 1) \) and
\[
x_\pm = \frac{3}{2}(1 + \sin^2 y \frac{\pi}{2}) \pm \sqrt{\frac{9}{4}(1 + \sin^2 y \frac{\pi}{2})^2 - \sin^2 y \frac{\pi}{2}} .
\]
It can be shown that
\[
0 < z_- \leq \frac{\sin^2 y \frac{\pi}{2}}{3} < 1
\]
and \( z_+ > 1 \). Therefore, \( \frac{\partial f}{\partial x} > 0 \) for \( x < x^* \) and \( \frac{\partial f}{\partial x} < 0 \) for \( x > x^* \), where \( x^* \) is the unique solution in \((0, 1)\) to
\[
\sin^2 x^* \frac{\pi}{2} = z_- .
\]
Therefore, \( f \) has a unique maximum in \((0, 1)\) at \( x^* \). Moreover, since for all \( x \in (0, 1) \),
\[
f(x, y) \leq F(\sin x \frac{\pi}{2}, \sin y \frac{\pi}{2}) ,
\]

17
where $F$ is defined by (6.9), we have

$$f(x^*, y) \leq \max_{0 < r < 1} F(r, \sin y \frac{\pi}{2}) = \frac{3}{4\sqrt{3}} \cot y \frac{\pi}{2}.$$  

We can now prove (a) in theorem 3.3:

We will only prove the inequality $\|\tilde{V}\|_1 \leq \sqrt{a} 0.733$. The proof of $\|\tilde{V}\|_\infty \leq \sqrt{\frac{1}{\alpha}} 0.733$ is completely analogous, by using (6.4) instead of (6.6).

By (6.6), we have

$$\|\tilde{V}\|_1 \equiv \max_{1 \leq j \leq n} \sum_{i=1}^{m_2} |v_{ij}| \leq \frac{\sqrt{a}}{2} \frac{1}{m_2 + 1} \max_{1 \leq j \leq n} \sum_{i=1}^{m_2} f(x_i, y_j).$$  

(6.13)

Let $h = \frac{1}{m_2 + 1} = x_1$. Since $\frac{d}{x_1(y_1)}$, $x_1(y_1) \in (0, 1)$ and, by lemma 6.2, $f$ is monotonic in the intervals $(0, x^*(y_1))$ and $(x^*(y_1), 1)$, it is easy to see, by using graphical arguments, that

$$h \sum_{i=1}^{m_2} f(x_i, y_j) \leq \int_{h}^{1} f(x, y_j) \, dx + hf(x^*(y_j), y_j),$$  

(6.14)

when $h \leq x^*(y_j)$. On the other hand, when $h > x^*(y_j)$, all the values of $x_i, i = 1, \ldots, m_2$ are on the interval $(x^*, 1)$, where $f$ is monotonically decreasing. Then, we have

$$h \sum_{i=1}^{m_2} f(x_i, y_j) = hf(h, y_j) + h \sum_{i=2}^{m_2} f(x_i, y_j) \leq \frac{1}{h} \int_{h}^{1} f(x, y_j) \, dx + hf(h, y_j).$$  

(6.15)

Let us first assume that $h \leq x^*(y_j)$. By (6.11) and (6.12), we can prove that $x^*(y) \leq y$ for all $y \in (0, 1)$. Then, by (6.7), we have:

$$\int_{h}^{1} f(x, y_j) \, dx \leq \frac{2}{2\pi} \left( \pi + g(h \frac{\pi}{2}, \sin y \frac{\pi}{2}) \right)$$  

(6.16)

because $g(z, w) \geq 0$ for $w < z$. Define the function $G(h, \beta) = g(h \frac{\pi}{2}, \sin \beta h \frac{\pi}{2})$. Then, the right hand side of (6.16) can be written as:

$$\frac{2}{2\pi} \left( \pi + G(h, \beta) \right),$$  

(6.17)

with $\beta = \frac{y_j}{h} \geq 1$. By differentiating $G$ with respect to $\beta$ we can see that $\frac{\partial G}{\partial \beta} < 0$ for all $h \in (0, 1)$ and $\beta \in [1, +\infty)$. Therefore, $G$ decreases with $\beta$, i.e.

$$G(h, \beta) \leq G(h, 1) = \lim_{w \to +} g(z, w) = \frac{1}{2} \log \frac{2 + \sqrt{2}}{2 - \sqrt{2}} - \frac{\pi}{2}.$$  

(6.18)

18
for all $h \in (0, 1)$ and $\beta \in [1, +\infty)$. We can then bound (6.16) by

$$
\int_{h}^{1} f(x, y_j) \, dx \leq \frac{1}{\sqrt{2\pi}} \left( \pi + \log \frac{2 + \sqrt{2}}{2 - \sqrt{2}} \right).
$$

(6.19)

On the other hand, by (6.10), we have

$$
h f(x^*(y_j), y_j) \leq \frac{3}{4\sqrt{3}} h \cot \beta h \frac{\pi}{2}.
$$

(6.20)

The right hand side of (6.20) can also be proven to decrease with $\beta$ and $h$ and therefore we have

$$
h f(x^*(y_j), y_j) \leq \frac{3}{4\sqrt{3}} h \cot \frac{\pi}{2} \leq \frac{3}{2\sqrt{3}\pi}.
$$

(6.21)

By replacing (6.19) and (6.21) in (6.14), we have

$$
h \sum_{i=1}^{m_2} f(x_i, y_j) \leq \frac{1}{\sqrt{2\pi}} \left( \pi + \log \frac{2 + \sqrt{2}}{2 - \sqrt{2}} \right) + \frac{3}{2\sqrt{3}\pi} = 1.4666
$$

(6.22)

and therefore, by (6.13), we have

$$
\|\tilde{V}\|_1 \leq \sqrt{\alpha} 0.7333
$$

(6.23)

when $h \leq x^*(y_j)$.

When $h > x^*(y_j)$, by (6.15) we have

$$
h \sum_{i=1}^{m_2} f(x_i, y_j) \leq \int_{x^*(y_j)}^{1} f(x, y_j) \, dx + h f(h, y_j).
$$

(6.24)

By (6.7),

$$
\int_{x^*(y_j)}^{1} f(x, y_j) \, dx \leq \frac{2}{\sqrt{2\pi}} \cos y_j \frac{\pi}{2} \left( \pi + g \left( \sin \frac{x^* \pi}{2}, \sin y_j \frac{\pi}{2} \right) \right)
$$

$$
\leq \frac{2}{\sqrt{2\pi}} \left( \pi + G(x^*, \frac{y_j}{x^*}) \right)
$$

and, by (6.18), we have

$$
\int_{x^*(y_j)}^{1} f(x, y_j) \, dx \leq \frac{1}{\sqrt{2\pi}} \left( \pi + \log \frac{2 + \sqrt{2}}{2 - \sqrt{2}} \right).
$$

(6.25)

Since $f(h, y_j) \leq f(x^*(y_j), y_j)$, by (6.21) we have

$$
h f(h, y_j) \leq \frac{3}{2\sqrt{3}\pi}
$$

(6.26)
By replacing (6.25) and (6.26) in (6.24), we have

\[ h \sum_{i=1}^{m_2} f(x_i, y_j) \leq 1.4666 \]

and therefore, by (6.13), we have

\[ \| \tilde{V} \|_1 \leq \sqrt{0.733} \]

when \( h > x^*(y_j) \). By (6.23) and (6.27), we proved that (6.23) holds for all \( h < 1 \).

**Proof of Theorem 4.1**

**Proof.** Define the function

\[ f(x) = \frac{1 + x - \sqrt{x}}{1 - x} . \]

We can easily prove that

\[ f(x) \geq 0.866 \quad \text{for all} \quad x \in [0, 1) . \]

By (4.4) and (2.12), we have

\[ \lambda_j(n, \beta_1, \beta_2) - |\delta_j(n, \beta_2)| = \left( \frac{1 + \gamma_j^{m_1+1}}{1 - \gamma_j^{m_1+1}} + \frac{1 + \gamma_j^{m_2+1}}{1 - \gamma_j^{m_2+1}} - 2 \frac{\gamma_j^{m_2+1}}{1 - \gamma_j^{m_2+1}} \right) \sqrt{\sigma_j + \frac{\sigma_j^2}{4}} . \]

Since \( \gamma_j < 1 \) and \( m_1 \leq m_2 \), we have

\[ \lambda_j(n, \beta_1, \beta_2) - |\delta_j(n, \beta_2)| \geq 2 \left( \frac{1 + \gamma_j^{m_2+1}}{1 - \gamma_j^{m_2+1}} - \frac{\gamma_j^{m_2+1}}{1 - \gamma_j^{m_2+1}} \right) \sqrt{\sigma_j + \frac{\sigma_j^2}{4}} 
= 2 \sqrt{\sigma_j + \frac{\sigma_j^2}{4}} f(\gamma_j^{m_2+1}) 
\geq 1.73 \sqrt{\sigma_j + \frac{\sigma_j^2}{4}} \]

Expressions for the elements of \( Q_{SE} \) and \( Q_{NE} \) of (4.11) are given in appendix A. We can easily verify that the elements of both matrices have the same absolute values. Also, both \( M_8 \) and \( M_9 \) have eigenvalues that are bounded from below by \( 4 \sin \frac{\pi}{2(m_1+1)} \).

By (4.10), (4.12) and (6.30), we can see that \( \| V \|_1 \) is bounded by

\[ \| V \|_1 \leq \frac{1}{\sqrt{0.866}} \left( \frac{\sqrt{\alpha}}{2} \frac{1}{m_1 + 1} \max_{1 \leq i \leq n} \sum_{i=1}^{m_1} \sqrt{\frac{\sin x_i \frac{\pi}{2}}{\sin^2 x_i + \sin^2 y_j \frac{\pi}{2}}} \sqrt{\frac{\cos x_i \frac{\pi}{2}}{\sin^2 x_i + \sin^2 y_j \frac{\pi}{2}}} \cos \frac{x_i \pi}{2} \sqrt{\sin y_j \frac{\pi}{2}} \cos \frac{y_j \pi}{2} \right) , \]

where \( x_i = i/(m_1 + 1) \) and \( y_j = j/(n + 1) \), for \( i = 1, \ldots, m_1 \) and \( j = 1, \ldots, n \). The proof of theorem 3.3 applies now to the expression in parenthesis.
References


