Advances in Applied Mathematics and Mechanics
Adv. Appl. Math. Mech., Vol. **6**, No. 3, pp. 327-344

DOI: 10.4208/aamm.2012.m69
June 2014

An Efficient Numerical Solution Method for Elliptic Problems in Divergence Form

Ali Abbas*

Department of Mathematics, Lebanese International University, Bekaa-Rayak, Lebanon

Received 28 December 2012; Accepted (in revised version) 19 December 2013

Available online 21 May 2014

Abstract. In this paper the problem $-\mathrm{div}(a(x,y)\nabla u)=f$ with Dirichlet boundary conditions on a square is solved iteratively with high accuracy for u and ∇u using a new scheme called "hermitian box-scheme". The design of the scheme is based on a "hermitian box", combining the approximation of the gradient by the fourth order hermitian derivative, with a conservative discrete formulation on boxes of length $2h$. The iterative technique is based on the repeated solution by a fast direct method of a discrete Poisson equation on a uniform rectangular mesh. The problem is suitably scaled before iteration. The numerical results obtained show the efficiency of the numerical scheme. This work is the extension to strongly elliptic problems of the hermitian box-scheme presented by Abbas and Croisille (J. Sci. Comput., 49 (2011), pp. 239–267).

AMS subject classifications: 65N35, 65N08

Key words: Hermitian scheme, box-scheme, Kronecker product, fast solver, iterative method, Poisson problem.

1 Introduction

Many references like Collatz [1], Forsythe and Wasow [2], Mitchell and Griffiths [3] and Iserles [4] treat the numerical resolution of partial differential equations as an educative building block in Applied Mathematics and Scientific Computing. For recent works, we refer to [5–11]. Beyond the design of specific numerical schemes which deals with accuracy and stability, the need of an efficient fast solver is a crucial issue to perform practical computations. The use of such solvers in canonical geometries remains at the heart of many computing codes in physics. Examples are among others fluid dymanics (compressible or incompressible Navier-Stokes equations), [12–14], the Helmholtz equation [15], computations in astrophysics, [16] or in geophysics, [17]. The scheme referred

*Corresponding author.
Email: ali.abbas@liu.edu.lb (A. Abbas)

as *hermitian box-scheme* was first introduced in [18]. It combines a finite volume "box" approach with a hermitian computation of the derivative. The practical resolution is performed by a direct resolution algorithm using the Sherman-Morrison formula based on the Fast Fourier Transform following previous works like [19], or [20] in a different context. This approach proves to be very efficient from the computing point of view. In [21] theoretical and numerical studies of the hermitian box-schemes using finite element methods is given. The main properties of the hermitian box-schemes compared to other methods like finite element method are the fourth order accuracy for u and ∇u on regular problems, very good capability to handle sharp contrast in the diffusion coefficients and a great flexibility in the design permitted by the variation of the quadrature rule for the gradient. In [22], a new hermitian box-scheme in one dimension (called HB-scheme) is introduced and analysed with approximations of order 1 of the derivatives on the boundary. This scheme is applied to solve regular elliptic problems and elliptic problems with high contrast in ellipticity. The rate of convergence varies between 1 and 2.5 according to the regularity of the problem. In [23], we have introduced a new fourth order compact scheme on a cartesian grid for the Poisson problem in a square, whose design is based on the preliminary work [22]. As the approximations of the derivatives on the boundary is raised to order three (instead of order 1 in [22]), the HB-scheme appears numerically to be fourth order accurate for u and ∇u.

We have also introduced a fast solver (called HB-solver) based on the Sherman-Morrisson formula and Fast Fourier Transform. It is proved that HB-solver is of complexity $\mathcal{O}(N^2\log_2(N))$, where N is the number of collocation points.

Our motivation is to use the HB-scheme and the HB-solver to solve more complicated problems. The problems that are considered are nonseparable elliptic problems in the form

$$\begin{cases} -\mathrm{div}(a(x,y)\nabla u)=f & \text{on } \Omega=(a,b)^2, \\ u=g & \text{on } \bar{\Omega}, \end{cases} \tag{1.1}$$

with Dirichlet boundary conditions. The outline of this paper is as follows. In Section 2, we give the notations and we descirbe the principles of the scheme on the 2D Poisson problem, then we extend this scheme to nonseparable elliptic problem (1.1). In Section 3, we present in details the basis of Concus-Golub's algorithm and the iterative procedure combined with the fast Poisson solver of [23]. In Section 4, we focus on numerical tests in 2D. We observe a remarkable superconvergence of the solution and its gradient for regular problems.

2 Principle of the Hermitian box-scheme in two dimensions

This section is devoted to the principle of the hermitian box-scheme (HB-Scheme) in two dimensions. We start by summarizing the finite difference and matrix notations, then we recall the matrix form of the HB-scheme for the Poisson problem in two dimensions [23] and we give the matrix form for nonseparable elliptic problems.

2.1 Finite difference operators and matrix notation

We lay out on $\Omega=(a,b)^2$ a regular grid

$$x_i=a+ih, \quad y_j=a+jh, \quad 0\le i,j\le N, \tag{2.1}$$

with step size $h=(b-a)/N$. The unknowns are located at discrete points $(x_i,y_j)_{0\le i,j\le N}$. The interior points are (x_i,y_j), $1\le i,j\le N-1$, and the boundary points are (x_i,y_j) with $i,j\in\{0,N\}$.

• *Laplacian.*

The two-dimenional Laplacian is

$$\delta_x^2 u_{i,j}=\frac{u_{i+1,j}+u_{i-1,j}-2u_{i,j}}{h^2}, \quad 1\le i\le N-1, \quad 0\le j\le N. \tag{2.2}$$

The Laplacian matrix $T\in\mathbb{M}_{N-1}(\mathbb{R})$ is

$$T=\begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & -1 & 2 & -1 \\ 0 & \dots & 0 & -1 & 2 \end{bmatrix}. \tag{2.3}$$

• *Simpson operator.*

The *Simpson finite-difference operator* σ_x is

$$\sigma_x u_{i,j}=\frac{1}{6}u_{i-1,j}+\frac{2}{3}u_{i,j}+\frac{1}{6}u_{i+1,j}, \quad 1\le i\le N-1, \quad 0\le j\le N. \tag{2.4}$$

Its matching matrix $P_s\in\mathbb{M}_{N-1}(\mathbb{R})$ is

$$P_s=I-T/6, \tag{2.5}$$

where I is the identity matrix of order $N-1$.

• *Centered difference operator.*

The centered operator δ_x is

$$\delta_x u_{i,j}=\frac{u_{i+1,j}-u_{i-1,j}}{2h}, \quad 1\le i\le N-1, \quad 0\le j\le N. \tag{2.6}$$

The matching matrix is the antisymmetric matrix $K\in\mathbb{M}_{N-1}(\mathbb{R})$ given by

$$K=\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & -1 & 0 & 1 \\ 0 & \dots & 0 & -1 & 0 \end{bmatrix}. \tag{2.7}$$

Similarly we define δ_y, δ_y^2, σ_y. These operators are defined at boundary points by 0. For example,

$$\sigma_x u_{0,j} = \delta_y u_{i,N} = 0, \quad 0 \le i,j \le N. \tag{2.8}$$

• Denoting $(e_i)_{1\le i\le N-1}$ the canonical basis of $\mathbb{R}^{N-1}$, the matrices $F_1, F_2 \in \mathbb{M}_{N-1}(\mathbb{R})$ are defined by

$$\begin{cases} F_1 = e_1 e_1^T + e_{N-1} e_{N-1}^T, \\ F_2 = -e_1 e_1^T + e_{N-1} e_{N-1}^T. \end{cases} \tag{2.9}$$

• Approximating the derivatives on the boundary points yield to the following matrices

$$\begin{cases} \mathcal{A} = -2F_2 + \dfrac{1}{2}(e_1 e_2^T - e_{N-1} e_{N-2}^T), \\ \mathcal{B} = 2F_1, \\ \mathcal{C} = \dfrac{5}{2} F_2. \end{cases} \tag{2.10}$$

2.2 Hermitian box-scheme for the Poisson problem in a square

Consider the Poisson problem on $\Omega = (a,b)^2$,

$$\begin{cases} -\Delta u = f & \text{on } \Omega, \\ u = g & \text{on } \partial\Omega. \end{cases} \tag{2.11}$$

We introduce the two variables $v^1 = \partial_x u$, $v^2 = \partial_y u$. The mixed form of the problem (2.11) is

$$\begin{cases} -\partial_x v^1 - \partial_y v^2 = f & \text{on } \Omega, \\ v^1 - \partial_x u = 0 & \text{on } \bar{\Omega}, \\ v^2 - \partial_y u = 0 & \text{on } \bar{\Omega}, \\ u = g & \text{on } \partial\Omega. \end{cases} \tag{2.12}$$

The domain Ω is discretized by (2.1). At each point of the grid are attached three unknowns, one for the function $u_{i,j} \simeq u(x_i,y_j)$ and two for the gradient, $u_{x,i,j} \simeq \partial_x u(x_i,y_j)$, $u_{y,i,j} \simeq \partial_y u(x_i,y_j)$. Then we have the unknowns $u = (u_{i,j})$, $u_x = (u_{x,i,j})$, $u_y = (u_{y,i,j})$. We call l_h^2 the space of grid functions $v = (v_{i,j})_{0\le i,j\le N}$. The square cells $k_{i,j}$ centered at (x_i,y_j) with length $2h$ are

$$K_{i,j} = [x_i - h, x_i + h] \times [y_j - h, y_j + h], \quad \forall 1 \le i,j \le N-1. \tag{2.13}$$

In addition, we define the operator vec_2 from l_h^2 into $\mathbb{R}^{(N-1)^2}$ by

$$\mathrm{vec}_2(v) = \sum_{i,j=1}^{N-1} (e_i \otimes e_j) v_{i,j}, \tag{2.14}$$

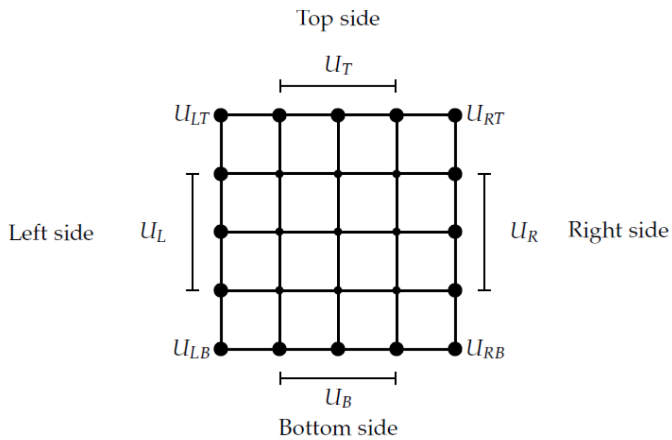


Figure 1: Notations corresponding to the domain $\Omega=(a,b)^2$. The interior points of discretization are represented by small circles. Big circles represent the points of discretization on $\partial\Omega$.

where $\otimes$ denotes the Kroneckerian product. For a summary and basic properties of the Kronecker product of matrices, we refer to [24]. For recent applications in fast computing in high dimensions [25, 26].

Actually, the operator vec_2 maps a matrix of order $N+1$ into a vector of order $(N-1)^2$ using the lexical order on i and j for the interior entries of the matrix. In the sequel, the small letters denote grid functions and big letters denote vectors in $\mathbb{R}^{(N-1)^2}$.

Therefore, the three discrete unknowns at internal points are

$$U=\text{vec}_2(u), \quad U_x=\text{vec}_2(u_x), \quad U_y=\text{vec}_2(u_y). \tag{2.15}$$

For example, for $N=3$ we have $u=(u_{i,j})_{0\le i,j\le 3}$ and $U=\text{vec}_2(u)=[u_{11},u_{12},u_{21},u_{22}]^T$.

Here we denote by $U_L,U_R\in\mathbb{R}^{N-1}$ the left and right Dirichlet boundary data at $x=a$ and $x=b$ respectively. Similarly, $U_B,U_T\in\mathbb{R}^{N-1}$ are the Bottom and Top Dirichlet data at $y=a$ and $y=b$ respectively,

$$\begin{cases} U_L=[u_{0,1},\cdots,u_{0,N-1}]^T, & U_R=[u_{N,1},\cdots,u_{N,N-1}]^T, \\ U_B=[u_{1,0},\cdots,u_{N-1,0}]^T, & U_T=[u_{1,N},\cdots,u_{N-1,N}]^T. \end{cases} \tag{2.16}$$

The boundary gradient vectors in $\mathbb{R}^{N-1}$ are denoted by $(U_{x,L},U_{y,L})$, $(U_{x,R},U_{y,R})$, $(U_{x,T},U_{y,T})$, and $(U_{x,B},U_{y,B})$. For example, the derivative vector with respect to x on the left side of Ω is

$$U_{x,L}=[u_{x,0,1},u_{x,0,2},\cdots,u_{x,0,N-2},u_{x,0,N-1}]^T. \tag{2.17}$$

We denote the four corner values $u_{0,0}$, $u_{0,N}$, $u_{N,0}$, $u_{N,N}$ by

$$U_{LB}=u_{0,0}, \quad U_{LT}=u_{0,N}, \quad U_{RB}=u_{N,0}, \quad U_{RT}=u_{N,N}. \tag{2.18}$$

Similar notations are used for the derivatives, for example:

$$(U_{x,LB},U_{y,LB})=(u_{x,0,0},u_{y,0,0}). \tag{2.19}$$

It has been proved in [23] that the algebraic structure of the HB-scheme can be interpreted in a simple way using Kronecker matrix algebra. In many situations one can take advantage of that structure to develop fast resolution procedures. The HB-scheme for the Poisson problem (2.11) has the following algebraic form: Find $U\in\mathbb{R}^{(N-1)^2}$ solution of

$$\frac{1}{h^2}(\mathcal{H}\otimes P_s+P_s\otimes\mathcal{H})U=F-G_x-G_y. \tag{2.20}$$

The two vectors G_x and $G_y\in\mathbb{R}^{(N-1)^2}$ are calculated in terms of Dirichlet boundary conditions:

$$\left\{\begin{aligned} G_x=&\frac{1}{h^2}(\mathcal{G}\otimes P_s)(e_1\otimes U_L+e_{N-1}\otimes U_R)\\ &+\frac{1}{6h^2}\big(\mathcal{H}U_B+\mathcal{G}(e_1U_{LB}+e_{N-1}U_{RB})\big)\otimes e_1\\ &+\frac{1}{6h^2}\big(\mathcal{H}U_T+\mathcal{G}(e_1U_{LT}+e_{N-1}U_{RT})\big)\otimes e_{N-1}, \end{aligned}\right. \tag{2.21}$$

and

$$\left\{\begin{aligned} G_y=&\frac{1}{h^2}(P_s\otimes\mathcal{G})(U_B\otimes e_1+U_T\otimes e_{N-1})\\ &+\frac{1}{6h^2}e_1\otimes\big(\mathcal{H}U_L+\mathcal{G}(e_1U_{LB}+e_{N-1}U_{LT})\big)\\ &+\frac{1}{6h^2}e_{N-1}\otimes\big(\mathcal{H}U_R+\mathcal{G}(e_1U_{RB}+e_{N-1}U_{RT})\big), \end{aligned}\right. \tag{2.22}$$

where the matrices $\mathcal{H}$ and $\mathcal{G}$ are given by

$$\left\{\begin{aligned} \mathcal{H}&=-\frac{1}{4}(K-F_2\mathcal{B})\Big(P_s-\frac{1}{6}\mathcal{B}\Big)^{-1}\Big(K-\frac{1}{3}\mathcal{A}\Big)-\frac{1}{2}F_2\mathcal{A},\\ \mathcal{G}&=-\frac{1}{4}(K-F_2\mathcal{B})\Big(P_s-\frac{1}{6}\mathcal{B}\Big)^{-1}\Big(F_2-\frac{1}{3}\mathcal{C}\Big)-\frac{1}{2}F_2\mathcal{C}. \end{aligned}\right. \tag{2.23}$$

The vector $F=\mathrm{vec}_2(\Pi^0 f_{i,j})\in\mathbb{R}^{(N-1)^2}$ is the second member vector with

$$\Pi^0 f_{i,j}=\frac{1}{4h^2}\int_{y_{j-1}}^{y_{j+1}}\int_{x_{i-1}}^{x_{i+1}}f(x,y)dxdy. \tag{2.24}$$

In practice $\Pi^0 f_{i,j}$ is approximated using the fourth-order Simpson formula

$$\left\{\begin{aligned} \Pi^0 f_{i,j}\simeq&\frac{1}{36}f_{i-1,j-1}+\frac{2}{18}f_{i-1,j}+\frac{1}{36}f_{i-1,j+1}\\ &+\frac{2}{18}f_{i,j-1}+\frac{4}{9}f_{i,j}+\frac{2}{18}f_{i,j+1}\\ &+\frac{1}{36}f_{i+1,j-1}+\frac{2}{18}f_{i+1,j}+\frac{1}{36}f_{i+1,j+1}. \end{aligned}\right. \tag{2.25}$$

In addition, we have proved that the derivatives vectors $U_x, U_y \in \mathbb{R}^{(N-1)^2}$ are expressed in terms of the solution vector $U \in \mathbb{R}^{(N-1)^2}$ and the boundary data by

$$\begin{cases} U_x = \dfrac{1}{h}\big((\mathcal{D}\otimes I)U + (\mathcal{E}\otimes I)(e_1\otimes U_L + e_{N-1}\otimes U_R)\big), \\ U_y = \dfrac{1}{h}\big((I\otimes\mathcal{D})U + (I\otimes\mathcal{E})(U_B\otimes e_1 + U_T\otimes e_{N-1})\big), \end{cases} \tag{2.26}$$

where the matrices $\mathcal{D}$, $\mathcal{E}$ are

$$\begin{cases} \mathcal{D} = \dfrac{1}{2}\Big(P_s - \dfrac{1}{6}\mathcal{B}\Big)^{-1}\Big(K - \dfrac{1}{3}\mathcal{A}\Big), \\ \mathcal{E} = \dfrac{1}{2}\Big(P_s - \dfrac{1}{6}\mathcal{B}\Big)^{-1}\Big(F_2 - \dfrac{1}{3}\mathcal{C}\Big). \end{cases} \tag{2.27}$$

2.3 Hermitian box-scheme for nonseparable elliptic problems

In fact, the idea of Concus and Golub [27] transforms the operator $-\mathrm{div}(a(x,y)\nabla u)$ into one whose differentiable part is $-\Delta$:

$$\begin{cases} (-\Delta + p)w = q & \text{on } \Omega, \\ w = a^{1/2}g & \text{on } \partial\Omega. \end{cases} \tag{2.28}$$

We introduce the two variables $v^1 = \partial_x w$ and $v^2 = \partial_y w$ then the mixed form of the problem (3.9) is

$$\begin{cases} -\partial_x v^1 - \partial_y v^2 + pw = q & \text{on } \Omega, \\ v^1 - \partial_x w = 0 & \text{on } \bar{\Omega}, \\ v^2 - \partial_y w = 0 & \text{on } \bar{\Omega}, \\ w = a^{1/2}g & \text{on } \partial\Omega. \end{cases} \tag{2.29}$$

Proposition 2.1. The hermitian box-scheme for the problem (3.9) can be written in matrix form as

$$\underbrace{\frac{1}{h^2}\Big(\mathcal{H}\otimes P_s + P_s\otimes\mathcal{H}\Big)W + \mathrm{vec}_2(\Pi^0(pw))}_{M^{\mathcal{H}}_{p,h}W} + G_x + G_y = Q, \tag{2.30}$$

with $W = \mathrm{vec}_2(w)$, $Q = \mathrm{vec}_2(\Pi^0 q) \in \mathbb{R}^{(N-1)^2}$ and $G_x, G_y \in \mathbb{R}^{(N-1)^2}$ are the vectors corresponding to the Dirichlet boundary conditions:

$$\begin{cases} G_x = \dfrac{1}{h^2}(\mathcal{G}\otimes P_s)(e_1\otimes W_L + e_{N-1}\otimes W_R) \\ \qquad\quad + \dfrac{1}{6h^2}\big(\mathcal{H}W_B + \mathcal{G}(e_1 W_{LB} + e_{N-1}W_{RB})\big)\otimes e_1 \\ \qquad\quad + \dfrac{1}{6h^2}\big(\mathcal{H}W_T + \mathcal{G}(e_1 W_{LT} + e_{N-1}W_{RT})\big)\otimes e_{N-1}, \end{cases} \tag{2.31}$$

and

$$\left\{\begin{aligned} G_y=&\frac{1}{h^2}(P_s\otimes\mathcal{G})(W_B\otimes e_1+W_T\otimes e_{N-1})\\ &+\frac{1}{6h^2}e_1\otimes\big(\mathcal{H}W_L+\mathcal{G}(e_1W_{LB}+e_{N-1}W_{LT})\big)\\ &+\frac{1}{6h^2}e_{N-1}\otimes\big(\mathcal{H}W_R+\mathcal{G}(e_1W_{RB}+e_{N-1}W_{RT})\big),\end{aligned}\right. \tag{2.32}$$

where the matrices P_s, $\mathcal{H}$ and $\mathcal{G}$ are given explicitly in (2.5) and (2.23).

Proof. The problem (3.9) rewrites

$$\left\{\begin{aligned} &-\Delta w=q-pw && \text{in } \Omega,\\ &w=a^{1/2}g && \text{on } \partial\Omega.\end{aligned}\right. \tag{2.33}$$

Now, we consider the problem (2.33) as a Poisson problem with second member $q-pw$. Therefore, using (2.20), the matrix form of (2.33) is found to be

$$\frac{1}{h^2}\Big(\mathcal{H}\otimes P_s+P_s\otimes\mathcal{H}\Big)W=\mathrm{vec}_2(\Pi^0(q-pw))-G_x-G_y. \tag{2.34}$$

Obviously

$$\mathrm{vec}_2(\Pi^0(q-pw))=\mathrm{vec}_2(\Pi^0 q)-\mathrm{vec}_2(\Pi^0(pw)). \tag{2.35}$$

Let $Q=\mathrm{vec}_2(\Pi^0 q)$, (2.34) becomes

$$\frac{1}{h^2}(\mathcal{H}\otimes P_s+P_s\otimes\mathcal{H})W=Q-\mathrm{vec}_2(\Pi^0(pw))-G_x-G_y, \tag{2.36}$$

where (2.30). □

Corollary 2.1. *The derivatives vectors* $W_x,W_y\in\mathbb{R}^{(N-1)^2}$ *of the problem* (3.9) *verify*

$$\left\{\begin{aligned} W_x=&\frac{1}{h}(\mathcal{D}\otimes I)W+\frac{1}{h}(\mathcal{E}\otimes I)(e_1\otimes W_L+e_{N-1}\otimes W_R),\\ W_y=&\frac{1}{h}(I\otimes\mathcal{D})W+\frac{1}{h}(I\otimes\mathcal{E})(W_B\otimes e_1+W_T\otimes e_{N-1}),\end{aligned}\right. \tag{2.37}$$

where $\mathcal{D}$ *and* $\mathcal{E}$ *are the matrices given in* (2.23).

Proof. By interchanging the roles of u and w, we remark that the second and the third equations of the mixed form (2.29) are exactly the same as the second and the third equations of (2.12). Therefore, we deduce from (2.26) the matrix form (2.37) of the derivatives W_x and W_y. □

Corollary 2.2. *Let $\mathcal{K}$ be a constant. By approximating the vector* $\mathrm{vec}_2(\Pi^0 w_{i,j})$ *as in* (2.25)*, the hermitian box-scheme of the problem*

$$\begin{cases} (-\Delta+\mathcal{K})w=q & \text{in } \Omega, \\ w=0 & \text{on } \partial\Omega, \end{cases} \tag{2.38}$$

has the following matrix form

$$\underbrace{\frac{1}{h^2}(\mathcal{H}\otimes P_s+P_s\otimes\mathcal{H})+\mathcal{K}(P_s\otimes P_s)}_{\Delta^{\mathcal{H}}_{\mathcal{K},h}W}=Q, \tag{2.39}$$

with $W=\mathrm{vec}_2(w)$, $Q=\mathrm{vec}_2(\Pi^0 q)\in\mathbb{R}^{(N-1)^2}$.

Proof. Let $p(x,y)=\mathcal{K}$ in (3.10). We deduce from the Proposition 2.1 that the problem (2.38) has the matrix form

$$\frac{1}{h^2}\big(\mathcal{H}\otimes P_s+P_s\otimes\mathcal{H}\big)W+\mathrm{vec}_2\big(\Pi^0(\mathcal{K}w)\big)+G_x+G_y=Q. \tag{2.40}$$

The homogeneous boundary conditions (2.38) leads to

$$G_x=G_y=0. \tag{2.41}$$

The same approximation as (2.25) gives

$$\begin{aligned}\Pi^0 w_{i,j}\simeq&\frac{1}{36}w_{i-1,j-1}+\frac{2}{18}w_{i-1,j}+\frac{1}{36}w_{i-1,j+1}\\ &+\frac{2}{18}w_{i,j-1}+\frac{4}{9}w_{i,j}+\frac{2}{18}w_{i,j+1}\\ &+\frac{1}{36}w_{i+1,j-1}+\frac{2}{18}w_{i+1,j}+\frac{1}{36}w_{i+1,j+1}.\end{aligned}$$

Using the homogeneous Dirichlet boundary conditions (2.38) we get $w_{i,j}=0$ for $i,j\in\{0,N\}$ and we can easily verify that

$$\Pi^0 w_{i,j}\simeq\big((P_s\otimes P_s)W\big)_{i,j}. \tag{2.42}$$

Let $W=\mathrm{vec}_2(w)$ then

$$\mathrm{vec}_2(\Pi^0 w)\simeq(P_s\otimes P_s)W. \tag{2.43}$$

Finally by using (2.43) in (2.40), we conclude (2.39). $\square$

3 The basis of Concus-Golub's algorithm

In its simplest form, the iterative procedure solves on a uniform square mesh the problem

$$\begin{cases} \mathcal{L}u=f & \text{on } \Omega, \\ u=g & \text{on } \partial\Omega. \end{cases} \tag{3.1}$$

In (3.1), $\Omega=(a,b)^2$ is a square and $\mathcal{L}$ represents the elliptic operator

$$\mathcal{L}u(x,y)=-\frac{\partial}{\partial_x}\Big(a(x,y)\frac{\partial u}{\partial x}(x,y)\Big)-\frac{\partial}{\partial_y}\Big(a(x,y)\frac{\partial u}{\partial y}(x,y)\Big). \tag{3.2}$$

We suppose that $a(x,y)$ is strictly positive on Ω and its boundary

$$0<a_{\min}\le a(x,y)\le a_{\max}. \tag{3.3}$$

There is also $\mathcal{L}_0=-\Delta$,

$$\mathcal{L}_0 u=-\frac{\partial^2 u}{\partial x^2}-\frac{\partial^2 u}{\partial y^2}. \tag{3.4}$$

We assume $a(x,y)$, $f(x,y)$ and $g(x,y)$ to be such that the solution is sufficiently well-behaved near the corner of Ω. The positivity of $a(x,y)$ implies that $\mathcal{L}$ is positive definite. We perform the following change of variable given by Concus and Golub [27],

$$w(x,y)=[a(x,y)]^{1/2}u(x,y). \tag{3.5}$$

We assume $a^{1/2}$ is twice differentiable. In this case, the first equation of (3.1) becomes

$$a^{-1/2}\mathcal{L}\equiv-\Delta w+pw=q, \tag{3.6}$$

with

$$\begin{cases} p=a^{-1/2}\Delta(a^{1/2}), \\ q=a^{-1/2}f. \end{cases} \tag{3.7}$$

The boundary conditions for $w(x,y)$ are

$$w=a^{1/2}g \quad \text{on } \partial\Omega. \tag{3.8}$$

Therefore, the problem (3.1) is equivalent to

$$\begin{cases} \mathcal{M}w=q & \text{on } \Omega, \\ w=a^{1/2}g & \text{on } \partial\Omega, \end{cases} \tag{3.9}$$

with

$$\mathcal{M} = -\Delta + p. \tag{3.10}$$

We propose an iterative method for the numerical resolution of problem (3.1) using the hermitian box-scheme (2.30). The problem (3.1) should be dimensionless form. The method due to Concus and Golub [27]. Among others, it uses the following relaxation principle. Suppose the original nonseparable problem is

$$\mathcal{L}u = f. \tag{3.11}$$

After change of variables, (3.11) rewrites

$$\mathcal{M}w = q. \tag{3.12}$$

In order to solve (3.12), we consider an iterative procedure for the elliptic equations of type (3.1) on a square. The method is based on the modified form of the iterative method (relaxation)

$$-\Delta u^{n+1} = -\Delta u^n - \tau(\mathcal{L}u^n - f), \tag{3.13}$$

where $\mathcal{L}u = f$ is the original problem to solve and τ is a parameter >0. The difficulty is that (3.13) can have a slow convergence in case as the spectral radius is close to 1 [27]. The main idea introduced by Concus and Golub consists to use a "shift" in (3.13). Consider the problem (3.9) (problem obtained after "scaling"). The iterative scheme used in this paper has the form

$$(-\Delta + \mathcal{K})w^{n+1} = (-\Delta + \mathcal{K})w^n - \tau(\mathcal{M}w^n - q), \tag{3.14}$$

where $\mathcal{K}$ is a constant, τ is a parameter >0 and Δ is the Laplace operator. The resolution of (3.14) is performed using a fast Poisson solver for $-\Delta + \mathcal{K}$. It is shown in [27] that if the classical second order five points Laplacian is used as the basic solver of $-\Delta + \mathcal{K}$ and if the constant $\mathcal{K}$ is given by

$$\mathcal{K} = \frac{1}{2}(\min p + \max p), \tag{3.15}$$

where $\min p$ and $\max q$ denote the minima and the maxima of $p(x,y)$ on the closed domain $\bar{\Omega}$, then the optimal choice of τ yielding the smallest spectral radius of the iteration matrix is

$$\tau = 1. \tag{3.16}$$

We present the resolution of problem (3.9) by the hermitian box-scheme (2.30) using the iterative method (3.14). The fast Poisson solver is exaclty the Algorithm 4.2 in [23]. In this work we focus on the iterative procedure of the scheme, we intend to study, in a second paper, the conditions yielding to the smallest spectral radius. The discrete operators corresponding to the continuous operator $\mathcal{M}$ and $-\Delta+\mathcal{K}$ are $M^{\mathcal{H}}_{p,h}$ in (2.30) and $\Delta^{\mathcal{H}}_{\mathcal{K},h}$ in (2.39) respectively. It follows from the Proposition 2.1 that the discrete form of the operator $\mathcal{M}w-q$ is

$$M^{\mathcal{H}}_{p,h}W+G_x+G_y-Q, \tag{3.17}$$

where $Q=\mathrm{vec}_2(\Pi^0 q)$. Notice that the presence of the operator $-\Delta+\mathcal{K}$ on both sides of (3.14) has the effect of cancelling the vectors of boundary conditions in discrete form. That is why we imposed homogeneous Dirichlet boundary conditions in (2.38). Furthermore, the discrete form of (3.14) is

$$\Delta^{\mathcal{H}}_{\mathcal{K},h}W^{n+1}=\Delta^{\mathcal{H}}_{\mathcal{K},h}W^{n}-\tau(M^{\mathcal{H}}_{p,h}W^n+G_x+G_y-Q). \tag{3.18}$$

Finally, (3.18) is an iterative method for the resolution of the elliptic problem (3.9).

4 Numerical results

In this section, we give some numerical results for the nonseparable problem (3.9) using the iterative method (3.18). First, we choose $\tau=1$, $\mathcal{K}=0$ and an initial vector $W^0\in\mathbb{R}^{(N-1)^2}$. Second, we perform the iteration: *For* $n=1,2,\cdots,iter$,

$$W^{n+1}=W^n-(\Delta^{\mathcal{H}}_{\mathcal{K},h})^{-1}(M^{\mathcal{H}}_{p,h}W^n+G_x+G_y-Q), \tag{4.1}$$

where *iter* is some fixed number of iterations. The code is performed in Matlab. The CPU time is computed by the function *tic* and *toc*. The domain Ω is the normalized square $(0,1)^2$. In the numerical tables u_{ex} is the exact solution and u is the computed solution. The boundary conditions are deduced by the exact solution. The average operator Π^0 defined in (2.24) is approximated by the (tensorial) Simpson formula in (2.25), which is fourth order. The errors are computed using the following L^2 and uniform norms

$$\begin{cases} \|u_{ex}-u\|_h=\Big(h^2\sum\limits_{i,j=1}^{N-1}(u_{ex}(x_i,y_j)-u_{i,j})^2\Big)^{\frac{1}{2}},\\ \|u_{ex}-u\|_\infty=\max\limits_{i,j=1,\cdots,N-1}|u_{ex}(x_i,y_j)-u_{i,j}|. \end{cases} \tag{4.2}$$

The rate of convergence is calculated by

$$\text{Convergence rate}=\log_2(e_{N/2}/e_N),$$

where e_N is the error obtained on a mesh of size $N\times N$. The stopping criteria used in all the following numerical tests is $\|\mathrm{U_{i+1}}-\mathrm{U_i}\|_2\le\mathrm{tol}$ where $\mathrm{tol}=10^{-6}$ and U_i,U_{i+1} are the computed solutions at the iterations i and $i+1$ respectively. Furthermore, The derivative vectors $W_x,W_y\in\mathbb{R}^{(N-1)^2}$ are computed by (2.37).

Table 1: Radius of convergence of the matrix of iteration if $p(x,y)=\alpha$, $\forall(x,y)\in\Omega$.

α	0	1/2	1	2
Radius of convergence	4.458(-15)	0.0253	0.0507	0.1013

In [22], many problems with discontinuous coefficients are tested using another quadrature formula near the point of discontinuity. In [27], the change of variable (3.5) is used at discrete level. Notice that the discrete transform used in [27] is based on the explicit form of the standard five-points Laplacian scheme and it is not possible with the proposed HB-scheme. In this section, we consider only cases where $a(x,y)$ is smooth. The scheme behaves very well, even on coarse grids, with very good error levels.

In order to see the behavior of the convergence of the algorithm we start by giving the table of radius convergence of the matrix of iteration

$$I-(\Delta^{\mathcal{H}}_{\mathcal{K},h})^{-1}(M^{\mathcal{H}}_{p,h}).$$

Suppose that $p(x,y)$ is constant i.e., $p(x,y)=\alpha$, $\forall(x,y)\in\Omega$ then $M^{\mathcal{H}}_{p,h}$ given in (2.30) becomes

$$M^{\mathcal{H}}_{p,h}=\frac{1}{h^2}(\mathcal{H}_2\otimes P_s+P_s\otimes\mathcal{H}_2)+\alpha(P_s\otimes P_s), \tag{4.3}$$

and the matrix of iteration becomes

$$I-\Big(\frac{1}{h^2}(\mathcal{H}_2\otimes P_s+P_s\otimes\mathcal{H}_2)^{-1})\Big)^{-1}\Big(\frac{1}{h^2}(\mathcal{H}_2\otimes P_s+P_s\otimes\mathcal{H}_2)+\alpha(P_s\otimes P_s)\Big).$$

In order to have an idea on the convergence of the algorithm, We give in Table 1 the spectral radius of the matrix of iteration for different values of the constant α. We remark that the spectral radius of this matrix is related to the values of α.

Test 4.1. In this test, we consider the exact solution $u_{ex}(x,y)=x+y$ and $a(x,y)=(1+x+y)^2$. We observe in Table 2 a convergence to the exact solution in only one iteration. In fact, this is due to the following. We have $p(x,y)=0$, then the matrix of iteration is reduced to

$$I-\Big(\frac{1}{h^2}(\mathcal{H}_2\otimes P_s)+(P_s\otimes\mathcal{H}_2)\Big)^{-1}\Big(\frac{1}{h^2}(\mathcal{H}_2\otimes P_s)+(P_s\otimes\mathcal{H}_2)\Big). \tag{4.4}$$

The radius of convergence of this matrix is equal to zero. This led to the convergence in one iteration as (4.1) becomes

$$W^{n+1}=(\Delta^{\mathcal{H}}_{\mathcal{K},h})^{-1}(G_x+G_y-Q), \tag{4.5}$$

which is a direct resolution of the Poisson problem (2.28) with $p(x,y)=0$.

Table 2: Error and convergence rate corresponding to Test 4.1.

$N_x\times N_y$	$\|u_{ex}-u\|_h$	$\|u_{x,ex}-u_x\|_h$	$\|u_{ex}-u\|_\infty$	$\|u_{x,ex}-u_x\|_\infty$	time(s.)	iterations
8×8	1.443(-14)	6.919(-14)	1.302(-13)	4.425(-13)	0.10	1
Convergence rate	1.69	5.47	0.66	3.41		
16×16	4.466(-15)	1.554(-15)	8.190(-14)	4.147(-14)	0.13	1
Convergence rate	-1.95	-8.22	-0.27	-6.68		
32×32	1.730(-14)	4.675(-13)	9.914(-14)	4.267(-12)	0.19	1
Convergence rate	0.20	-0.67	-0.23	-1.13		
64×64	1.999(-14)	7.473(-13)	1.167(-13)	9.353(-12)	0.34	1

Table 3: Error and convergence rate corresponding to Test 4.2.

$N_x\times N_y$	$\|u_{ex}-u\|_h$	$\|u_{x,ex}-u_x\|_h$	$\|u_{ex}-u\|_\infty$	$\|u_{x,ex}-u_x\|_\infty$	time(s.)	iterations
8×8	1.055(-12)	2.943(-12)	2.163(-12)	6.003(-12)	0.34	13
Convergence rate	1.98	1.89	1.98	1.88		
16×16	2.664(-13)	7.940(-13)	5.478(-13)	1.624(-12)	0.58	13
Convergence rate	-0.50	-0.55	-0.50	-0.82		
32×32	3.793(-13)	1.164(-12)	7.789(-13)	2.872(-12)	1.03	13
Convergence rate	1.97	1.68	1.95	0.12		
64×64	9.646(-14)	3.614(-13)	2.004(-13)	2.629(-12)	2.54	13

Table 4: Error and convergence rate corresponding to Remark 4.1.

$N_x\times N_y$	$\|u_{ex}-u\|_h$	$\|u_{x,ex}-u_x\|_h$	$\|u_{ex}-u\|_\infty$	$\|u_{x,ex}-u_x\|_\infty$	time(s.)	iterations
16×16	4.484(-7)	3.843(-6)	2.302(-6)	2.191(-5)	0.22	5
Convergence rate	3.97	3.32	3.79	2.70		
32×32	3.083(-8)	3.851(-7)	1.653(-7)	3.370(-6)	0.39	5
Convergence rate	3.99	3.39	3.89	2.78		
64×64	1.939(-9)	3.662(-8)	1.113(-8)	4.901(-7)	0.91	5
Convergence rate	4.02	3.44	3.94	2.88		
128×128	1.192(-10)	3.374(-9)	7.239(-10)	6.657(-8)	2.52	5
Convergence rate	4.06	3.46	3.97	2.93		
256×256	7.122(-12)	3.062(-10)	4.616(-11)	8.686(-9)	8.21	5

Test 4.2. The exact solution is

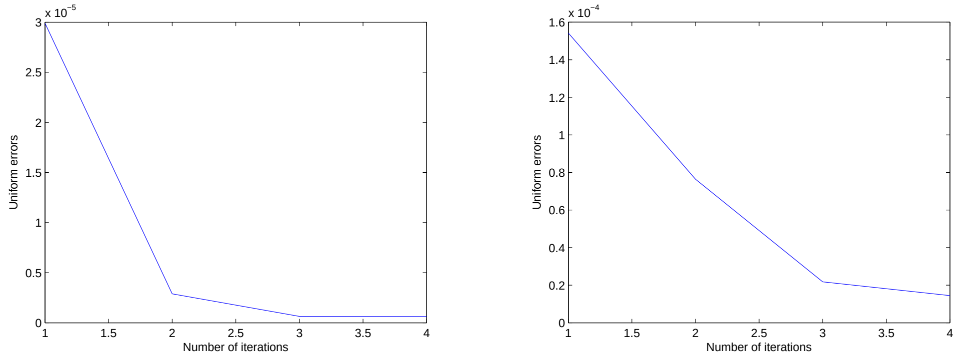
$$u_{ex}(x,y)=\frac{1}{2}\Big[\Big(x-\frac{1}{2}\Big)^2+\Big(y-\frac{1}{2}\Big)^2\Big]$$

with

$$a(x,y)=\Big[1+\frac{1}{2}\sin\Big(\frac{\pi}{2}(x+y)\Big)\Big]^2.$$

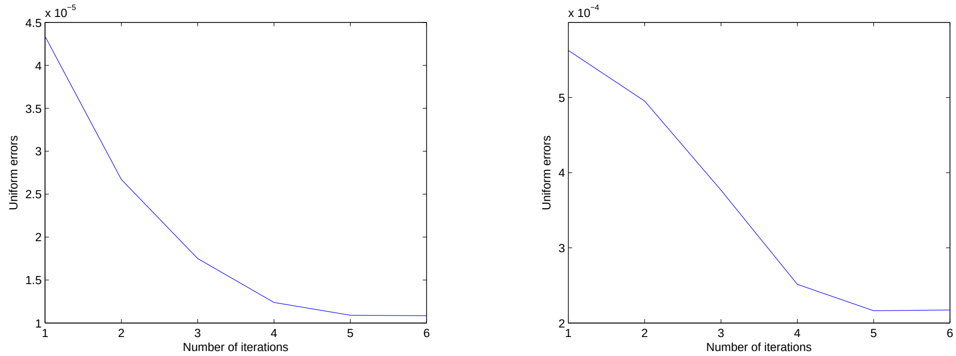
The numerical results given in Table 3 show a convergence to the exact solution in 13 iterations. The computer accuracy is reached in this test.

Test 4.3. We suppose $u_{ex}(x,y)=x^2+y^2$ and $a(x,y)=\delta+x+y$. We use two different values of δ. We observe in Table 4 for $\delta=1$ an error equal to $4.616(-11)$ for $N=256$ comparing to $5.608(-8)$ for $\delta=0.1$. This lack of accuracy is due to the oscillation of the exact solution and its derivatives which are monitored by the parameter δ.



(a) (b)

Figure 2: Uniform errors vs. Number of iterations of Remark 4.2 for the solution (a) and its derivative (b).



(a) (b)

Figure 3: Uniform errors vs. Number of iterations of Test 4.4 for the solution (a) and its derivative (b).

Remark 4.1. We suppose $\delta=1$. The numerical results are given in Table 4. The convergence rate is 4 for the solution and it converges to 3 for the gradient in the uniform norm (4.2).

Remark 4.2. We suppose $\delta=0.1$. For this value of δ, the solution and the derivatives are more oscillating. The uniform errors vs. the number of iterations for the solution and its derivative are given in Fig. 2. The errors are less accurate than those obtained for $\delta=1$.

However, in Remarks 4.1 and 4.2, a very good accuracy is attained on the final grid 256×256 for both the exact solution and its gradient.

Test 4.4. Oscillating diffusion coefficient.
In this example, we consider a diffusion coefficient in the form

$$a(x,y)=(4+\sin(4\pi x)\sin(4\pi y))^2, \tag{4.6}$$

with the exact solution $u_{ex}(x,y)=\sin(2\pi x)\sin(2\pi y)$.

This case is more difficult because $a(x,y)$ and its derivatives are oscilliant. We observe in Fig. 3 the uniform errors vs. the number of iterations for the solution and its derivative. We refer to [28] for more comments on the difficulty of this case. The results obtained in [28] are more accurate. However, here the derivatives are computed, which is not the case in [28].

Test 4.5. Functions of class C^2 and C^3.
The Simpson formula (2.6) requires functions of class C^4 to provide good approximations. We have observed numerically that the proposed scheme fails to converges for many tests where the solution is of class C^2. In this test, we consider the following exact solution which is of class C^3 on $[0,1]^2$,

$$u_{ex}(x,y)=(x+y-1)^{\alpha}, \tag{4.7}$$

where $\alpha=18/5$ and $a(x,y)=1+x+y$. We observe in Table 5 the good level errors of the scheme.

Table 5: Error and convergence rate corresponding to Test 4.5.

$N_x\times N_y$	$\|u_{ex}-u\|_h$	$\|u_{x,ex}-u_x\|_h$	$\|u_{ex}-u\|_\infty$	$\|u_{x,ex}-u_x\|_\infty$	time(s.)	iterations
32×32	9.737(-7)	2.000(-6)	5.573(-6)	2.951(-5)	1.21	10
Convergence rate	4.00	3.89	3.3.47	3.03		
64×64	6.023(-8)	1.349(-7)	4.911(-7)	3.603(-6)	2.82	10
Convergence rate	4.00	3.85	3.49	3.00		
128×128	3.744(-9)	9.352(-9)	4.368(-8)	4.469(-7)	8.21	10
Convergence rate	4.00	3.82	3.49	3.00		
256×256	2.333(-10)	6.587(-10)	3.870(-9)	5.574(-8)	24.76	10

5 Conclusions

This paper introduces a methodology to design compact finite difference schemes in cartesian geometries. It also introduces an iterative method with a new scheme to solve strongly elliptic problems in divergence form. The scheme does not use any kind of staggered grid. All the unknowns are located at points (x_i,y_j), [23,29]. The iterative method is based on the repeated solution by a fast direct method of a discrete Poisson equation on a rectangular mesh. The numerical results show the effectiveness of the scheme and the iterative technique as well. The computational cost is $k(\mathcal{O}(N^2\log_2(N)))$ where k is the number of iterations and N is the number of collocation points.

The work is going on in several directions, including the theoretical study of the spectral radius of the matrix of iteration. In addition, a strategy to generalize this scheme to irregular geometries using embedded grids. As in previous work [30], our approach will use finite-volume discretizations which embeds the domain in a regular cartesian grid and treat the solution as cell centered on a rectangular grid even when the cell centers are outside the domain. In order to conserve the fourth order accuracy of the scheme, we will use fourth order accurate fluxes on each cell that contains a portion of the boundary.

Acknowledgments

The author expresses his gratitude to the anonymous referees for their constructive critics which helped greatly to enhance this paper. Thanks also to Prof. Jean-Pierre Croisille and Prof. Raafat Talhouk for their encouragements and advice.

References

[1] L. COLLATZ, The Numerical Treatment of Differential Equations, 3-rd edition, Springer-Verlag, 1960.

[2] G. E. FORSYTHE AND W. R. WASOW, Finite Difference Methods for Partial Differential Equations 6th edition, John Wiley & Sons, Applied Mathematics Series, 1960.

[3] A. R. MITCHELL AND D. F. GRIFFITHS, The Finite Difference Method in Partial Differential Equations, John Wiley & Sons, 1980.

[4] A. ISERLES, A First Course in the Numerical Analysis of Differential Equations, Cambridge Univ. Press, 1996.

[5] G. COPPOLA AND C. MEOLA, *Generalization of the spline interpolation based on the principle of compact schemes*, J. Sci. Comput., 2002.

[6] S. K. LELE, *Compact finite-difference schemes with spectral-like resolution*, J. Comput. Phys., 1992.

[7] K. MAHESH, *A family of high order finite difference schemes with good spectral resolution*, J. Comput. Phys., 1998.

[8] T. K. SENGUPTA, G. GANERIWAL AND D. DE, *Analysis of central and upwind compact schemes*, J. Comput. Phys., 2003.

[9] S. E. SHERER AND J. N. SCOTT, *High order compact finite difference methods on general overset grids*, J. Comput. Phys., 2005.

[10] YIN WANG AND JUN ZHANG, *Sixth order compact scheme combined with multigrid method and extrapolation technique for 2D Poisson equation*, J. Comput. Phys., 2009.

[11] W. HUNDSDORFER AND J. G. VERWER, *Numerical solution of time-dependent advection-diffusion-reaction equations*, Springer Series in Computational Mathematics, Springer, 2010.

[12] Y. MORINISHI, T. S. LUND, O. V. VASILYEV AND P. MOIN, *Fully conservative higher order finite difference schemes for incompressible flows*, J. Comput. Phys., 1998.

[13] G. H. GOLUB, L. C. HUANG, H. SIMON AND W-P. TANG, *A fast Poisson solver for the finite difference solution of the incompressible Navier-Stokes equations*, SIAM J. Sci. Comput., 1998.

[14] M. BEN-ARTZI, J-P. CROISILLE AND D. FISHELOV, Navier-Stokes Equations in Planar Domains, Imperial College Press, 2012.

[15] E. BRAVERMAN, M. ISRAELI AND A. AVERBUCH, *A hierarchical 3-D direct Helmholtz solver by domain decomposition and modified Fourier method*, SIAM J. Sci. Comput., 2005.

[16] P. LONDRILLO, *Adaptive grid-based gas-dynamics and Poisson solvers for gravitating systems*, Mem. A. A. It. Suppl., 2004.

[17] K. SHIRAISHI AND T. MATSUOKA, *Wave propagation simulation using the CIP method of characteristics equations*, Commun. Comput. Phys., 2008.

[18] H. B. KELLER, *A new difference scheme for parabolic problems*, Numerical Solution of Partial Differential Equations, II, Academic Press, New York, 1971.

[19] P. BJORSTAD, *Fast numerical solution of the biharmonic Dirichlet problem on rectangles*, SIAM J. Numer. Anal., 1983.

[20] M. Ben-Artzi, J-P. Croisille and D. Fishelov, *A fast direct solver for the biharmonic problem in a rectangular grid*, SIAM J. Sci. Comput., 2008.
[21] I. Greff, Schémas boîte: Etude Théorique et Numérique, PhD thesis, Université de Metz, France, 2003.
[22] J-P. Croisille, *A Hermitian box-scheme for one-dimensional elliptic equations-application to problems with high contrasts in the ellipticity*, Computing, 2006.
[23] A. Abbas and J-P. Croisille, *A fourth order Hermitian Box-Scheme with fast solver for the Poisson problem in a square*, J. Sci. Comput., 2011.
[24] D. A. Harville, Matrix Algebra from a Statistician Perspective, Springer, 2008.
[25] L. Grasedyck, *Existence and computation of low Kronecker-Rank approximations for large linear systems of tensor product structure*, Computing, 2004.
[26] W. Hackbusch, B. N. Khoromskij and E. E. Tyrtyshnikov, *Hierarchical Kronecker tensor-product approximations*, J. Numer. Math., 2005.
[27] P. Concus and G. H. Golub, *Use of fast direct methods for the efficient numerical solution of nonseparable elliptic equations*, SIAM J. Numer. Anal., 1973.
[28] E. Braverman, B. Epstein, M. Israeli and A. Averbuch, *A fast spectral subtractional solver for elliptic equations*, J. Sci. Comput., 2004.
[29] A. Abbas, *A fourth order Hermitian box-scheme with fast solver for the Poisson problem in a cube*, Numer. Methods Partial Differential Equation, doi:10.1002/num.21807, 2013.
[30] H. Johansen and P. Colella, *A cartesian grid embedded boundary method for Poisson's equation on irregular domains*, J. Comput. Phys., 1998.