Stabilized Times Schemes for High Accurate Finite Differences Solutions of Nonlinear Parabolic Equations

November 3, 2016

Matthieu Brachet[1] and Jean-Paul Chehab[2]

[1]Institut Elie Cartan de Lorraine, Université de Lorraine,
Site de Metz, Bât. A Ile du Saucy, F-57045 Metz Cedex 1

[2]Laboratoire Amienois de Mathématiques Fondamentales et Appliquées (LAMFA), UMR 7352
Université de Picardie Jules Verne, 33 rue Saint Leu, 80039 Amiens France

Abstract

The Residual Smoothing Scheme (RSS) have been introduced in [1] as a backward Euler's method with a simplified implicit part for the solution of parabolic problems. RSS have stability properties comparable to those of semi-implicit schemes while giving possibilities for reducing the computational cost. A similar approach was introduced independently in [11, 12] but from the Fourier point of view. We present here a unified framework for these schemes and propose practical implementations and extensions of the RSS schemes for the long time simulation of nonlinear parabolic problems when discretized by using high order finite differences compact schemes. Stability results are presented in the linear and the nonlinear case. Numerical simulations of 2D incompressible Navier-Stokes equations are given for illustrating the robustness of the method.

Keywords: Preconditioning, Stability, Finite Differences, Compact Schemes, Times Schemes, Navier-Stokes Equations

AMS Classification[2010]: 65F08,65M06,65M12,76D05

1 Introduction

It is a common fact in numerical analysis that the choice of a time marching scheme must balance stability, accuracy and reasonable computational cost. Typically, when considering e.g. the numerical solution of a space-discretized parabolic problems, such as

$$\begin{aligned} &\frac{du}{dt} + Au = f, \\ &u(0) = u_0, \end{aligned} \tag{1}$$

where A denotes the stiffness matrix, it is well known that the implicit schemes are stable but need an additional problem to be solved at each step while the explicit schemes are very cheap but suffer of a hard time step limitation making them bad suited for capturing the long time behavior of the solutions. An ideal scheme should combine stability and low computational cost (explicity) for a comparable accuracy.

Two independent attempts have been made successfully in that direction :

First, A. Cohen and *al* proposed in [1] the following stabilization to the forward Euler scheme (Residual Smoothing Scheme or RSS) for the discretized parabolic equations associated to homogeneous Dirichlet (or Neumann) Boundary conditions:

$$\frac{u^{(k+1)}-u^{(k)}}{\Delta t}+ \quad \underbrace{\tau\, B(u^{(k+1)}-u^{(k)})}_{\text{Stabilization term}} \quad +Au^{(k)}=f, \tag{2}$$

where τ is a positive real number to be chosen and B a preconditioner of the stiffness matrix A. Originally introduced in the context of wavelet discretizations, the matrix B can be taken as the diagonal part of A and is then an inconditional preconditioner: the new scheme is no more expensive than the classical forward Euler's while the stability is increased. However, RSS is only first order accurate in time and, in order to increase the accuracy, it was proposed in [1] to apply a Richardson extrapolation; typically a second order of accuracy was obtained as shown by numerical evidences. A rough analysis of the stabilized and extrapolated scheme was made by Ribot and Schatzman [20, 21], they derived stability and error estimates in energy norms .

Independently, Costa, Dettori, Gottlieb and Temam have introduced in [11, 12] a similar approach but starting from a Fourier-analysis point of view, in the context of multiresolution methods of nonlinear Galerkin type for spectral discretizations (Fourier, Chebyshev), see [23]. They proposed to stabilize the forward Euler scheme for the heat equation

$$u_t - u_{xx} = f,$$

by adding a stabilization term of the form $\beta\dfrac{du}{dt}$. The new scheme writes in Fourier basis as

$$\frac{\hat{u}_m^{(k+1)}-\hat{u}_m^{(k)}}{\Delta t} \quad +\underbrace{\frac{\beta_j}{\Delta t}(\hat{u}_m^{(k+1)}-\hat{u}_m^{(k)})}_{\text{Stabilization term}} \quad = -m^2\hat{u}_m^{(k)}+\hat{f}_m,\ m=N_{j-1},\cdots,N_j, \tag{3}$$

where we have decomposed the frequency range $[1,N]$ into $\cup_{j=1}^{d}[N_{j-1},N_j]$. In other words

$$\frac{\hat{u}^{(k+1)}-\hat{u}^{(k)}}{\Delta t}-B(\hat{u}^{(k+1)}-\hat{u}^{(k)})=-D\hat{u}^{(k)}+\hat{f},$$

with $D=diag(1,4,9,\cdots N^2)$ and

$$B=\left(\begin{array}{cccc} B_1 & 0 & & 0 \\ 0 & \ddots & & \\ & & \ddots & \\ 0 & & & B_d \end{array}\right),$$

with $B_j = \beta_j Id_j$, Id_j being the identity matrix of size $(N_j - N_{j-1}) \times (N_j - N_{j-1})$ (we have set $N_0 = 0$ and $N_d = N$); B is then a preconditioner of D in the Fourier space. In the linear case, these two approaches coincide. Of course, same framework can be derived when considering orthogonal polynomials. Costa and Chehab [7, 8] have extended this scheme to hierarchical discretizations in finite differences.

The main advantage of the RSS approach is that a simplified (yet costless) solver is used for the implicit part of the time marching scheme while displaying comparable stability properties to Backward's Euler Scheme. One situation of particular interest, on which we focus in the present work, occurs when handling high order discretizations of the stiffness matrix A, e.g. with finite differences compact schemes. In that case, A is full, this is due to the implicit part of the scheme. Hence, matrix-vector product are costlty and must be reduced as far as possible. A lower level of space dicretization (say second order) generates a sparse stiffness matrix B which is a natural efficient preconditioner of A. Then, the RSS scheme can be implemented efficiently taking advantage of the existing (sometimes fast) solvers of the system of the form

$$(Id + \Delta t B)u = f,$$

such as sparse factorizations and FFT.
The RSS approach can be proposed to solve fully discretized time dependent PDEs with high accurate spatial discretization, with compact scheme, while using the computational facilities of the sparse numerical solvers (Fast solvers, limited memory).

In this article, we propose a unified approach to RSS-like schemes that rely [1] and [12]. We derive stability results in the linear and the nonlinear case; we also present practical and efficient adaptations for the high accurate finite differences solutions of nonlinear parabolic equations.

The paper is organized as follows: in Section 2, we derive a general approach to RSS schemes and we give stability results in the linear and the nonlinear case, the accuracy in time of the new scheme is also discuted. After that, in section 3, we describe the compact scheme discretization and the preconditioning that will be used. Section 4 is devoted to numerical illustrations, we compare RSS approach to the classical one with emphasis on the stability, the accuracy (particularly the dynamics to steady state) and for that purspose, we solve high accurate finite difference steady state of 2D incompressible Navier-Stokes equations (Lid driven cavity) and recover the results of the litterature.

2 Derivation of the stabilized schemes - properties

2.1 Formal Derivation of the stabilized schemes

Let us consider the finite dimensional differential system

$$\frac{du}{dt} + F(u) = 0, t > 0, \tag{4}$$

$$u(0) = u_0, \tag{5}$$

here $F : I\!R^N \to I\!R^N$ is a regular map. The backward Euler scheme applied to the above system generates the iterations

$$u^{(k+1)} - u^{(k)} + \Delta t F(u^{(k+1)}) = 0,$$

and a (possibly) nonlinear problem must be solved at each step. Making the approximation

$$F(u^{(k+1)}) \simeq F(u^{(k)}) + F'(u^{(k)})(u^{(k+1)} - u^{(k)}),$$

where $F'(u^{(k)})$ denotes the differential of F at $u^{(k)}$, we obtain the scheme

$$\frac{u^{(k+1)} - u^{(k)}}{\Delta t} + F'(u^{(k)})(u^{(k+1)} - u^{(k)}) + F(u^{(k)}) = 0,$$

so

$$u^{(k+1)} = u^{(k)} - \Delta t(Id + \Delta t F'(u^{(k)}))^{-1} F(u^{(k)}).$$

Setting $\Phi(v) = v - u^{(k)} + \Delta t F(v)$, we see that $u^{(k+1)}$ is nothing else but the first iteration of the Newton-Raphson scheme applied to $\Phi(v)$ when starting from the initial guess $u^{(k)}$.

Now, if we replace $F'(u^{(k)})$ by a preconditioner τB_k, we find

$$\frac{u^{(k+1)} - u^{(k)}}{\Delta t} + \quad \tau \underbrace{B_k(u^{(k+1)} - u^{(k)})}_{\text{Global stabilization}} \quad + F(u^{(k)}) = 0, \tag{6}$$

and $u^{(k+1)}$ is thus the first iteration of a quasi Newton Method applied to $\Phi(v)$ when starting from the initial guess $u^{(k)}$.

The efficiency of this stabilized scheme is closely related to the cost of the computation of the preconditioner of the jacobian matrix which changes at each iteration: technique of existing updating factorizations as those presented in [6] and [2] could be adapted.

In the present work, we will not discuss on the analysis of the nonlinear version of the scheme, say (7), but we will present on Section 4 numerical results obtained with this scheme. We will rather consider the semi linear approach: if $F(u)$ can be expressed as $F(u) = Au + f(u)$, we define teh scheme

$$\frac{u^{(k+1)} - u^{(k)}}{\Delta t} + \quad \tau \underbrace{B(u^{(k+1)} - u^{(k)})}_{\text{Stabilization of the linear part}} \qquad + F(u^{(k)}) = 0, \tag{7}$$

where B is a preconditioner of A.

It is important to point out that (7) is consistant with the computation of steady states and can then be applied as a pseudo-time numerical solver, as illustrated in Section 4 with impressible NSE.

Of course this stabilization approach applies to the linear case ($f(u) = 0$). Particularly, RSS is a simplified θ-scheme in which the matrix A is replaced by a preconditioner. Indeed, the θ-scheme write, after simplifications as

$$u^{(k+1)} = u^{(k+1)} - \Delta t(Id + \theta \Delta t A)^{-1}(Au^{(k)} - f),$$

and, substituting A by B in the implicit part, we recover the RSS

$$u^{(k+1)} = u^{(k+1)} - \Delta t(Id + \theta \Delta t B)^{-1}(Au^{(k)} - f),$$

with $\theta = \tau$.

In practice, the use of a preconditioner B of A leads to propose $K = Id + \tau\Delta t B$ as preconditioner of $M = Id + \Delta t A$, where Id is the identity matrix. This can be realized in many ways, e.g., by computing K as an incomplete factorization of M; in some cases it can be done by solving the linear systems involving K with fast solvers (FFT or so), see section 3. The RSS approach applies also to linear problems with a matrix $A(t)$ which depends on time t:

$$\frac{\partial u}{\partial t} + A(t)u = f, \tag{8}$$

that we discretize as

$$\underbrace{(Id + \tau\Delta t B_k)}_{K_k} \quad (u^{(k+1)} - u^{(k)}) \quad = F - A(k\Delta t)u^{(k)\cdot} \tag{9}$$

The matrix K_k can be computed as an incomplete LU factorization of $M = Id + \Delta t A(k\Delta t)$ and, if $A(t)$ does vary slightly with t, incremental factorization updates from K_{k-1} can be done following the techniques proposed in [6]. Notice also that scheme (10) can be obtained by applying RSS to linearized equation, as

$$\underbrace{(Id + \tau\Delta t B_k)}_{K_k} \quad (u^{(k+1)} - u^{(k)}) \quad = -\Delta t F(u^{(k)}), \tag{10}$$

where B_k is here such that $F(u^k) = B_k u^k$.

2.2 Properties of the schemes

2.2.1 The linear case

Let A and B be both $N \times N$ real symmetric positive definite matrices; the symmetry of A is considered for the sake of simplicity however the following approach remains valuable in the nonsymmetric case, see section 3 and Theorem 3.2. At this point we state the hypothesis ($\mathcal{H}$) on which we will base number of results along this work: we assume that there exist two strictly positive real numbers α and β such that

$$(\mathcal{H}) \qquad \alpha < Bu, u > \leq < Au, u > \leq \beta < Bu, u >, \ \forall u \in \mathbb{R}^N.$$

It is important to note that α and β can depend on the dimension N, if not the matrix B is said to be an inconditional preconditioner of A. We will use the following notations: $< .,. >$ is the euclidian scalar product in $\mathbb{R}^N$ and $\| \, . \, \|$, the associated norm. We will note λ_{min} (resp. λ_{max}) the lowest (resp. the largest) eigenvalue of A.

We now consider the RSS scheme applied to the discretized heat equation

$$\frac{u^{(k+1)} - u^{(k)}}{\Delta t} + \tau B(u^{(k+1)} - u^{(k)}) = -Au^{(k)}.$$

We first prove a simple stability result:

Proposition 2.1 *Under hypothesis $\mathcal{H}$, we have the following stability conditions:*

- *If $\tau \geq \frac{\beta}{2}$, the scheme is unconditionally stable (i.e. stable $\forall \; \Delta t > 0$)*
- *If $\tau < \frac{\beta}{2}$, then the scheme is stable for $0 < \Delta t < \frac{2}{\left(1 - \frac{2\tau}{\beta}\right) \rho(A)}$.*

Proof. Taking the usual scalar product of each terms with $u^{(k+1)} - u^{(k)}$, we find

$$\frac{1}{\Delta t} \| \, u^{(k+1)} - u^{(k)} \, \|^2 + \tau < B(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} >= - < Au^{(k)}, u^{(k+1)} - u^{(k)} > .$$

Using the parallelogram identity,

$$< Au^{(k)}, u^{(k+1)} - u^{(k)} >= \frac{1}{2} \left(< Au^{(k+1)}, u^{(k+1)} > - < A(u^{(k)}, u^{(k)} + < A(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)}) > \right),$$

we infer

$$\frac{1}{\Delta t} \| \, u^{(k+1)} - u^{(k)} \, \|^2 + \tau < B(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} >$$
$$-\tfrac{1}{2} \left(< Au^{(k)}, u^{(k)} > - < Au^{(k+1)}, u^{(k+1)} > + < A(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > \right) \;\; = 0.$$

Hence the stability condition $< Au^{(k)}, u^{(k)} > > < Au^{(k+1)}, u^{(k+1)} >$ holds when

$$\frac{1}{\Delta t} \| \, u^{(k+1)} - u^{(k)} \, \|^2 + \tau < B(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > - \frac{1}{2} < A(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} >> 0.$$

We have, using $\mathcal{H}$

$$\begin{gathered} \tau < B(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > -\tfrac{1}{2} < A(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > \\ \geq \\ \left(\tfrac{\tau}{\beta} - \tfrac{1}{2}\right) < A(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > (\geq 0). \end{gathered}$$

A sufficient stability condition is then

$$\frac{1}{\Delta t} \| \, u^{(k+1)} - u^{(k)} \, \|^2 + \left(\frac{\tau}{\beta} - \frac{1}{2}\right) < A(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} >> 0.$$

This is satisfied once $\dfrac{1}{\Delta t} + \lambda_{min}(\dfrac{\tau}{\beta} - \dfrac{1}{2}) > 0$. Therefore, if $\dfrac{\tau}{\beta} - \dfrac{1}{2} \geq 0$ the previous inequality holds for all $\Delta t > 0$, this means the stability $\forall \Delta t > 0$.

Now if $\tau < \dfrac{\beta}{2}$, then, since $< A(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > \leq \rho(A) \parallel u^{(k+1)} - u^{(k)} \parallel^2$, a sufficient condition of stability is

$$\frac{1}{\Delta t} + \left(\frac{\tau}{\beta} - \frac{1}{2}\right)\rho(A) > 0,$$

from which we deduce

$$0 < \Delta t < \frac{2}{\left(1 - 2\frac{\tau}{\beta}\right)\rho(A)},$$

as a sufficient stability condition. ■

We point out that if $B = A$ (then $\alpha = \beta = 1$) and $\tau = \theta \in [0, 1]$, the stability condition coincide with the one of the θ-scheme.

The stability is easily obtained when taking τ large enough. However, a too large value of τ deteriorates the consistency of the scheme and, as a particular effect, the convergence to the steady state is longer in time. In fact both the value of τ and the preconditioning quality of B act on the accuracy of the RRS scheme which remains always first order accurate in time as illustrated in Figure (1).

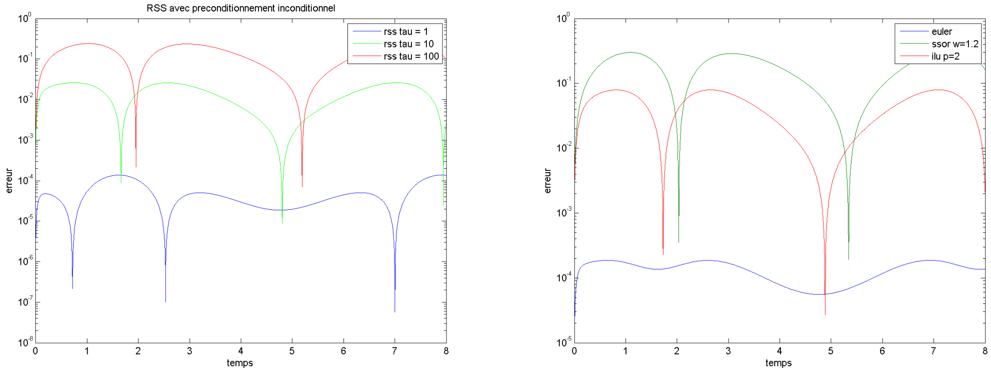

Figure 1: (left) Influence of the parameter τ on the accuracy of RSS. Error vs time for different values of τ - $N = 127$, $\Delta t = 0.004$, (right) Error vs time with ILU(2) and SSOR preconditioners, $N = 63$,$\tau 1$, $\Delta t = 0.004$

A first natural question is the choice of the best value τ_{opt} of τ, for a fixed preconditioner B; τ_{opt} can be simply computed such as minimizing $\parallel \tau B - A \parallel_F$. We easily show that

$$\tau_{opt} = \frac{trace(B^T A)}{\parallel B \parallel_F^2}.$$

We remark that when $B = Id$, then $\tau_{opt} = \frac{trace(A)}{n}$, the mean value of the eigenvalues of A.

A second natural question deals with the gain of stability brought by the RSS scheme as respect to an explicit method, namely forward Euler's for which the time step must be taken strictly lower than $\Delta t_c = \frac{2}{\rho(A)}$. In other words, for a given preconditioner matrix B and for a given number τ, we look to $\kappa > 1$ such that RSS is stable with time step $\Delta t = \kappa \Delta t_c$. We deduce directly from the previous computations.

Proposition 2.2
- *If $\tau \geq \frac{\beta}{2}$, RSS is infinitely more stable than backward Euler's.*
- *If $\tau < \frac{\beta}{2}$, then RSS is at least $\kappa = \frac{1}{1 - \frac{2\tau}{\beta}}$ times more stable than Euler's.*

Proof. We deduce from the proposition 2.2.1 that $\frac{2}{\left(1 - \frac{2\tau}{\beta}\right)\rho(A)} = \kappa \frac{2}{\rho(A)}$, then $\kappa = \frac{1}{\left(1 - \frac{2\tau}{\beta}\right)}$.
■

We now propose to quantify the consistency error with τ by comparison with the Backward Euler Scheme (which is unconditionally stable). Particularly we analyse the behavior of the difference of the sequences generated by the two schemes: the stabilization has as an effect to slow down the convergence in time to the steady state.

Proposition 2.3 *We consider the two sequences*

$$\frac{u^{(k+1)} - u^{(k)}}{\Delta t} + \tau B(u^{(k+1)} - u^{(k)}) = f - Au^{(k)},$$

and

$$\frac{v^{(k+1)} - v^{(k)}}{\Delta t} + Av^{(k+1)} = f,$$

with $u^{(0)} = v^{(0)}$. We let $M = Id - \Delta t(Id + \tau\Delta t B)^{-1}A$ and we assume that $\| M \| < 1$, then, there exists $\gamma \in [0, 1[$ such that

$$\| u^{(k)} - v^{(k)} \| \leq \Delta t^2 \| \tau B - A \| \frac{1}{1-\gamma} \| f - Av^{(0)} \|, \forall k \geq 0.$$

Proof. We first remark that $\| M \| < 1$ implies the stability of the RSS scheme, since M is the iteration matrix and $\rho(M) \leq \| M \|$.

We take the difference and we let $w^{(k)} = u^{(k)} - v^{(k)}$. We have

$$\frac{w^{(k+1)} - w^{(k)}}{\Delta t} + \tau B(w^{(k+1)} - w^{(k)}) + (\tau B - A)(v^{(k+1)} - v^{(k)}) = -Aw^{(k)}$$

Hence, after the usual simplifications, we can write

$$w^{(k+1)} = (Id - \Delta t(Id + \tau\Delta t B)^{-1}A)w^{(k)} - \Delta t(Id + \tau\Delta t B)^{-1}(\tau B - A)(v^{(k+1)} - v^{(k)}).$$

Using the definition of M, we obtain directly the estimate

$$\| w^{(k+1)} \| \leq \| M \| \| w^{(k+1)} \| + \Delta t \| (Id + \tau \Delta t B)^{-1} \| \| \tau B - A \| \| v^{(k+1)} - v^{(k)} \| .$$

We first remark that we have the relations

$$v^{(k+1)} - v^{(k)} = \Delta t (Id + \Delta t A)^{-1} (f - A v^{(k)}),$$

and

$$r^{(k+1)} = (Id - \Delta t A)(Id + \Delta t A)^{-1} r^{(k)},$$

where we have set $r^{(k)} = f - A v^{(k)}$. It follows that

$$v^{(k+1)} - v^{(k)} = \Delta t (Id + \Delta t A)^{-1} \left((Id - \Delta t A)(Id + \Delta t A)^{-1} \right)^k r^{(0)}.$$

Then

$$\| v^{(k+1)} - v^{(k)} \| \leq \Delta t \| (Id + \Delta t A)^{-1} \| \| (Id - \Delta t A)(Id + \Delta t A)^{-1} \|^k \| r^{(0)} \| .$$

We set $\gamma = \| (Id - \Delta t A)(Id + \Delta t A)^{-1} \|$, we have of course $\gamma < 1, \forall \Delta t > 0$. A simple induction gives

$$\| w^{(k+1)} \| \leq \| M \|^k \| w^{(0)} \| + \Delta t^2 \| (Id + \tau \Delta t B)^{-1} \| \| \tau B - A \| \| (Id + \Delta t A)^{-1} \| \sum_{j=0}^{k} \gamma^{(j)} \| M \|^{k-j} \| r^{(0)} \| .$$

Using $\| M \| < 1$ and $w^{(0)} = 0$ we find

$$\| w^{(k+1)} \| \leq + \Delta t \| (Id + \tau \Delta t B)^{-1} \| \| \tau B - A \| \Delta t \| (Id + \Delta t A)^{-1} \| \sum_{j=0}^{k} \gamma^{(j)} \| r^{(0)} \| .$$

Hence

$$\| w^{(k+1)} \| \leq \Delta t^2 \| \tau B - A \| \frac{1}{1 - \gamma} \| r^{(0)} \|,$$

which shows the dependence on $\| \tau B - A \|$. Of course if $\tau = 1$ and $B = A$ we have $w^{(k)} = 0, \forall k$. Hence the result. ■

We deduce immediately the

Corollary 2.4 *The RSS method is first order accurate.*

Proof. It suffices to own that the Backward Euler method is first order accurate. ■

We now will consider the particular the case in which B is a diagonal matrix. This choice allows a fast solution of the implicit part of RSS, the matrix $Id + \tau \Delta t B$ being also diagonal. In general situations, a diagonal preconditioner is not the most efficient, but in some cases, e.g. when the discrete problem in written in hierarchical-like bases, it is particularly interesting to consider B as a diagonal matrix: A. Cohen *et al* [1] have introduced the RSS scheme for a problem discretized in a wavelet basis, in which the diagonal part of the stiffness matrix is an inconditional preconditioner; this was independently applied by Costa *et al* [11, 12] using Fourier and Chebyshev expansions and then in finite differences with incrementals unkowns by Chehab and Costa [7, 8, 9]. The underlying idea is to decompose the unknowns U of the nodal basis into a hierarchy of arrays of details at different levels $(\hat{U}_0, \hat{U}_1, \cdots, \hat{U}_d)^T$; here $\hat{U}_0$ is associated to a coarse discretization and then captures only low frequencies while the other set of components $\hat{U}_j,\ j = 1, \cdots, d$ are details associated to refinment of the coarse approximation space and capture high frequencies.

The time limitation of the explicit schemes for parabolic problems depends on its capability to contain high frequencies expansions. In Fourier-like basis, the details attached to high frequencies are small quantities regardless to the details attached to low frequencies since they contribute residually to the energy norm of the signal. As proposed in [12, 7, 8] this situation allows to damp differently the high and the low frequencies components without deterioring the consistency of the scheme.

Consider for instance the heat equation

$$u_t - u_{xx} = f,$$

that we discretize in time with the forward Euler scheme as

$$\frac{u^{(k+1)} - u^{(k)}}{\Delta t} - u_{xx}^{(k)} = f.$$

In [11, 12], this scheme was proposed to be stabilized as

$$\frac{u^{(k+1)} - u^{(k)}}{\Delta t} + \beta \frac{u^{(k+1)} - u^{(k)}}{\Delta t} - u_{xx}^{(k)} = f.$$

Considering a Fourier discretization, we find, after simplifications

$$\frac{\hat{u}_m^{(k+1)} - \hat{u}_m^{(k)}}{\Delta t} - \frac{\beta}{\Delta t}(\hat{u}_m^{(k+1)} - \hat{u}_m^{(k)}) = -k^2 \hat{u}_m^{(k)} + \hat{f}_m,\ m = 1, \cdots, N.$$

If we decompose the frequency range $[1, N]$ into $\cup_{j=1}^d [N_{j-1}, N_j]$, the above scheme can be applied to each range with a different stabilizing parameter β_j, so we obtain, for $j = 1, \cdots, d$

$$\frac{\hat{u}_m^{(k+1)} - \hat{u}_m^{(k)}}{\Delta t} - \frac{\beta_j}{\Delta t}(\hat{u}_m^{(k+1)} - \hat{u}_m^{(k)}) = -m^2 \hat{u}_m^{(k)} + \hat{f}_m,\ m = N_{j-1}, \cdots, N_j.$$

Letting $\tilde{\beta}_i = \dfrac{\beta_i}{\Delta t}$, the last scheme is rewritten as

$$\hat{u}_m^{(k+1)} = \left(1 - \frac{m^2 \Delta t}{1 + \Delta t \tilde{\beta}_i}\right) \hat{u}_m^{(k)} + \frac{\Delta t}{1 + \Delta t \tilde{\beta}_i} \hat{f}_m,\ m = N_{j-1}, \cdots, N_j.$$

The stability condition is then

$$(m^2 - 2\tilde{\beta}_j)\Delta t < 2, m = N_{j-1}, \cdots, N_j.$$

The scheme is unconditionally stable at level j if $\tilde{\beta}_j \geq \frac{N_{j+1}^2}{2}$ and stable under condition $0 < \Delta t < \frac{2}{N_{j+1}^2 - 2\tilde{\beta}_j}$ otherwise. This condition is of course similar to that found in Proposition 2.2.1.

Now, considering all the components, we write

$$\frac{\hat{u}^{(k+1)} - \hat{u}^{(k)}}{\Delta t} - B(\hat{u}^{(k+1)} - \hat{u}^{(k)}) = -D\hat{u}^{(k)} + \hat{f},$$

with $D = diag(1, 4, 9, \cdots N^2)$ and

$$B = \begin{pmatrix} B_1 & 0 & & 0 \\ 0 & \ddots & & \\ & & \ddots & \\ 0 & & & B_d \end{pmatrix},$$

where $B_j = \beta_j Id_j$, Id_j being the identity matrix of size $(N_j - N_{j-1}) \times (N_j - N_{j-1})$ (we have set $N_0 = 0$ and $N_d = N$).

B is then a preconditioner of D in the Fourier space; in the linear case, these two approaches coincide. Costa and Chehab [7, 8] have extended this scheme to hierarchical discretizations in finite differences.
Note that this framework allows to damp high frequencies and to leave unchanged the low ones changing only slightly the consistency of the scheme while increasing its stability. This can be done typically by taking $\beta_j = 0$ on the low freqencies components and $\beta_j >> 1$ on the high ones.

As stated in the introduction, we concentrate on problems discretized in the nodal basis, however, it is useful to make the link with the hierarchical approach. We give a block version of Proposition 2.2.1.

We assume that the stiffness matrix A written in a detail basis posseses the following block decomposition

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{d,d} \\ A_{2,1} & & & \\ \vdots & & & \\ A_{d,1} & & \cdots & A_{d,d} \end{pmatrix}.$$

We note the corresponding block decomposition of a vector U as $U = (u_1, \cdots, u_d)^T$. We have the

Theorem 2.5 *A sufficient stability condition is*

$$\frac{1}{\Delta t} + \tau\beta_i - \frac{1}{4}\left(\sum_{j=1}^{d} \parallel A_{i,j} \parallel + \parallel A_{j,i} \parallel\right) > 0, i = 1, \cdots, d.$$

Proof. Taking the scalar product of the equation with $U^{(k+1)} - U^{(k)}$, we find

$$\frac{1}{\Delta t} \parallel U^{(k+1)} - U^{(k)} \parallel^2 + \tau < B(U^{(k+1)} - U^{(k)}), U^{(k+1)} - U^{(k)} >$$

$$-\frac{1}{2}\left(< AU^{(k)}, U^{(k)} > - < AU^{(k+1)}, U^{(k+1)} > + < A(U^{(k+1)} - U^{(k)}), U^{(k+1)} - U^{(k)} >\right) \quad = 0,$$

and using the block decomposition

$$\sum_{i=1}^{d} \frac{1}{\Delta t} \parallel u_i^{(k+1)} - u_i^{(k)} \parallel^2 + \tau\beta_i \parallel u_i^{(k+1)} - u_i^{(k)} \parallel^2$$

$$-\tfrac{1}{2}\left(< AU^{(k)}, U^{(k)} > - < AU^{(k+1)}, U^{(k+1)} >\right) + -\tfrac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} < A_{i,j}(u_j^{(k+1)} - u_j^{(k)}), u_i^{(k+1)} - u_i^{(k)} > \quad = 0,$$

A sufficient condition for the energy stability is then

$$\sum_{i=1}^{d} \frac{1}{\Delta t} \parallel u_i^{(k+1)} - u_i^{(k)} \parallel^2 + \tau\beta_i \parallel u_i^{(k+1)} - u_i^{(k)} \parallel^2 - \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} < A_{i,j}(u_j^{(k+1)} - u_j^{(k)}), u_i^{(k+1)} - u_i^{(k)} > > 0.$$

Since,

$$\begin{aligned}
-\tfrac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} < A_{i,j}(u_j^{(k+1)} - u_j^{(k)}), u_i^{(k+1)} - u_i^{(k)} > \quad & \geq -\tfrac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \parallel A_{i,j} \parallel | u_j^{(k+1)} - u_j^{(k)} || u_i^{(k+1)} - u_i^{(k)} | \\
& \geq -\tfrac{1}{4}\sum_{i=1}^{d}\sum_{j=1}^{d} \parallel A_{i,j} \parallel | u_j^{(k+1)} - u_j^{(k)} |^2 \\
& -\tfrac{1}{4}\sum_{i=1}^{d}\sum_{j=1}^{d} \parallel A_{i,j} \parallel | u_i^{(k+1)} - u_i^{(k)} |^2 \\
& = -\tfrac{1}{4}\sum_{i=1}^{d}\left(\sum_{j=1}^{d} \parallel A_{i,j} \parallel + \parallel A_{j,i} \parallel\right)
\end{aligned}$$

We find as sufficient stability condition

$$\sum_{i=1}^{d}\left(\frac{1}{\Delta t} + \tau\beta_i - \frac{1}{4}\left(\sum_{j=1}^{d} \parallel A_{i,j} \parallel + \parallel A_{i,j} \parallel\right)\right) \parallel u_i^{(k+1)} - u_i^{(k)} \parallel^2 > 0.$$

In particular, if $\tau\beta_i \geq \frac{1}{4}\left(\sum_{j=1}^{d} \parallel A_{i,j} \parallel + \parallel A_{i,j} \parallel\right) > 0$, the stability is unconditional. ■

It has to be noted that when $d = 1$ and $B = Id$, we recover the result given by Proposition 2.2.1.

We infer also that, the extradiagonal coefficients of stiffness matrices in hierarchical-like basis enjoy of a descreasing magnitude properties far from the diagonal making successful the approach.

If we consider Fourier basis, the stifness matrix A is diagonal and the stability condition at range j is

$$(m^2 - 2\tilde{\beta}_j)\Delta t < 2, m = N_{j-1}, \cdots, N_j$$

so, taking $\beta_j = \dfrac{N_{j+1}^2}{2}$, we have an incondiditionally stable RSS scheme.

2.2.2 The nonlinear case

We now aim at applying the RSS Scheme to reaction-diffusion equation say

$$\frac{\partial u}{\partial t} - \Delta u + \frac{1}{\epsilon^2} f(u) = 0, \quad x \in \Omega, t > 0, \tag{11}$$

$$\frac{\partial u}{\partial n} = 0 \qquad \partial\Omega, t > 0, \tag{12}$$

$$u(x,0) = u_0(x) \qquad x \in \Omega, \tag{13}$$

where $\epsilon > 0$ is a given parameter. The RSS scheme applied to the discretized scheme writes as

$$\frac{u^{(k+1)} - u^{(k)}}{\Delta t} + \tau B(u^{(k+1)} - u^{(k)}) + \frac{1}{\epsilon^2} f(u^{(k)}) = -Au^{(k)}. \tag{14}$$

We set $E(u) = \frac{1}{2} < Au, u > + \frac{1}{\epsilon^2} < F(u), 1 >$, where F is a primitive of f that we choose such that $F(0) = 0$. We say that the scheme is energy decreasing if

$$E(u^{(k+1)}) < E(u^{(k)})$$

Particularly, if F has nonnegative values, the scheme will be stable for the norm $\mid u \mid_A = \sqrt{< Au, u >}$. It is well known that treating the nonlinear term explicitely leads to an important time step reduction: typically the stability condition is $0 < \Delta t < C\epsilon^2$ where C is a constant that depends on f. Using standard computations we can prove that the following stability results for the RSS scheme:

Theorem 2.6 *Assume that f is $\mathcal{C}^1$ and $\mid f' \mid_\infty \leq L$. We have the following stability conditions*

- *If $\tau \geq \frac{\beta}{2}$ then*
  - *if $\left(\frac{\tau}{\beta} - \frac{1}{2}\right)\lambda_{min} - \frac{L}{2\epsilon^2} \geq 0$ then the scheme is unconditionally stable*
  - *if $\left(\frac{\tau}{\beta} - \frac{1}{2}\right)\lambda_{min} - \frac{L}{2\epsilon^2} < 0$ then the scheme is stable for*

$$0 < \Delta t < \frac{1}{\frac{L}{2\epsilon^2} - \left(\frac{\tau}{\beta} - \frac{1}{2}\right)\lambda_{min}}$$

- *If $\tau < \frac{\beta}{2}$ then the scheme is stable for*

$$0 < \Delta t < \frac{1}{\frac{L}{2\epsilon^2} - \left(\frac{\tau}{\beta} - \frac{1}{2}\right)\rho(A)}$$

Remark 2.7 *Notice that Shen et al [22] have proposed the scheme*

$$\frac{u^{(k+1)}-u^{(k)}}{\Delta t}+\frac{S}{\epsilon^2}(u^{(k+1)}-u^{(k)})+Au^{(k+1)}+\frac{1}{\epsilon^2}f(u^{(k)})=0. \tag{15}$$

With Theorem 2.2.2, we recover the stability conditions proposed by J. Shen when $A=B$ and $\tau=1$ and $S=0$; The term $\frac{S}{\epsilon^2}(u^{(k+1)}-u^{(k)})$ plays the role of the stabilizator, following the same principle as the schemes introduced in [11, 12]. The time restriction become harder when ϵ takes small values. This situation motivates the use of stabilized schemes. Now, if $S>\frac{L}{2}$, the scheme (15) is unconditionally stable. This is to be compared with the RSS scheme (14). We find as inconditional stability condition

$$\tau>\frac{\beta L}{2\lambda_{min}}+\frac{\beta\epsilon^2}{2},$$

which is a comparable condition for ϵ small enough and β bounded, since in pratice λ_{min} is a positive constant which not depend on the dimension of the problem, the first eigenvalue of the stifness matrix is indeed nicely captured by the discretization schemes. However the additional stabilizing terms can deteriorate the consistency.

The time stability condition is a real limitation for small values of ϵ and in addition the assumption $\mid f'\mid_\infty\leq L$ is restrictive. Also, we rather propose to adapt RSS-schemes to inconditionally stable schemes for Allen-Cahn equations (and without any limitation on the growth of f), that we present hereafter. These schemes need to solve at each step a nonlinear system of equations and the RSS approach will allows to simplify their solution when using a fixed point method. We indroduce the new schemes taking advantage of the analysis presented in [13, 14].

A first inconditionally stable scheme is ([**?**])

$$\frac{u^{(k+1)}-u^{(k)}}{\Delta t}+Au^{(k+1)}+\frac{1}{\epsilon^2}DF(u^{(k)},u^{(k+1)})=0 \tag{16}$$

where

$$DF(u,v)=\begin{cases}\frac{F(u)-F(v)}{u-v} & \text{if } u\neq v\\ f(u) & \text{if } u=v\end{cases}$$

The proof is obtained by taking the scalar product with $\frac{u^{(k+1)}-u^{(k)}}{\Delta t}$, using also the identity $<DF(u,v),(u-v)>=<F(u),1>-<F(v),1>$.

We can now consider the associated RSS-scheme

$$\frac{u^{(k+1)}-u^{(k)}}{\Delta t}+\tau B(u^{(k+1)}-u^{(k)})+DF(u^{(k+1)},u^{(k)})=-Au^{(k)} \tag{17}$$

We can prove the following result:

Proposition 2.8 *Under hypothsesis $\mathcal{H}$*

- *if* $\tau \geq \frac{\beta}{2}$, *the RSS scheme is unconditionally stable,*
- *if* $\tau < \frac{\beta}{2}$, *the RSS scheme is stable under condition*

$$0 < \Delta t < \frac{\beta}{\rho(A)(\frac{\beta}{2} - \tau)}.$$

Proof. We take the scalar product of (17) with $\frac{u^{(k+1)} - u^{(k)}}{\Delta t}$ and we find

$$\begin{array}{l} \| \frac{u^{(k+1)} - u^{(k)}}{\Delta t} \|^2 + \frac{\tau}{\Delta t} < B(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > \\ + \frac{1}{\Delta t} < DF(u^{(k+1)}, u^{(k)}), u^{(k+1)} - u^{(k)} > \end{array} = - < Au^{(k)}, u^{(k+1)} - u^{(k)} >$$

Hence, since

$$< Au^{(k)}, u^{(k+1)} - u^{(k)} > = \frac{1}{2} \Big(< Au^{(k)}, u^{(k)} > - < Au^{(k+1)}, u^{(k+1)} > + < A(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > \Big),$$

we obtain

$$\begin{array}{l} \| \frac{u^{(k+1)} - u^{(k)}}{\Delta t} \|^2 + \frac{1}{\Delta t} (\tau < B(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > \\ - \frac{1}{2} < A(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} >) + \frac{1}{\Delta t} \left(E(u^{(k+1)}) - E(u^{(k)}) \right) \end{array} = 0$$

It follows immediately that if

$$\frac{\|u^{(k+1)} - u^{(k)}\|^2}{\Delta t} \left(\frac{1}{\Delta t} + \tau \frac{< B(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} >}{\|u^{(k+1)} - u^{(k)}\|^2} - \frac{1}{2} \frac{< A(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} >}{\|u^{(k+1)} - u^{(k)}\|^2} \right) \geq 0,$$

then $E(u^{(k+1)}) \leq E(u^{(k)})$ and the scheme is stable. This condition is satisfied once

$$\frac{1}{\Delta t} + \left(\frac{\tau}{\beta} - \frac{1}{2} \right) \frac{< A(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} >}{\|u^{(k+1)} - u^{(k)}\|^2} \geq 0.$$

So, if $\tau \geq \frac{\beta}{2}$ the stability holds for every $\Delta t > 0$. If $\tau < \frac{\beta}{2}$, a suficient stability condition is

$$\frac{1}{\Delta t} + \left(\frac{\tau}{\beta} - \frac{1}{2} \right) \rho(A) > 0.$$

So

$$0 < \Delta t < \frac{\beta}{\rho(A)(\frac{\beta}{2} - \tau)}.$$

■

Remark 2.9 *Both schemes (16) and (17) need a nonlinear system of equation to be solved at each step, e.g., using a fixed point method. However, the one corresponding to (17) is simpler since A (which can be full) can be replaced by a simpler matrix B (eventually sparse) in the fixed point iterations.*

We now present an other RSS-scheme based on the so-called convex-splitting technique [14]. More generally, we can rewrite the PDE problem as

$$\frac{\partial u}{\partial t} - \Delta u + \nabla F(u) = 0 \quad x \in \Omega, t > 0 \tag{18}$$

$$\frac{\partial u}{\partial n} = 0 \quad x \in \partial\Omega, t > 0 \tag{19}$$

$$u(x,0) = u_0(x) \qquad x\Omega \tag{20}$$

We follow [14] and make the assumptions

$$\begin{array}{ll} F(u) \geq 0, & \forall u \in \mathbb{R}^n \\ F(u) \rightarrow +\infty & \text{as } \|u\| \rightarrow +\infty \\ < J(\nabla F)(u)u, u > \geq \lambda & \forall u \in \mathbb{R}^n \end{array} \tag{21}$$

We now make the following additional hypothesis $\mathcal{C}\wr\backslash\sqsubseteq$

i. We assume that F can be splitted as follows

$$F(u) = F_c(u) - F_e(u),$$

where $F_* \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$

ii. F_* is strictly convex in $\mathbb{R}^n$, $* = c$ or $* = e$.

iii. $< [\nabla F_e(u)]u, u > \geq -\lambda$, $\forall u \in \mathbb{R}^n$

We recall the following result, see [14] and the references therein

Lemma 2.10 *Assume that $F*$ satisfies (21), then, there exists $\lambda \in \mathbb{R}$ such that*

$$F(u) - F(v) \leq < \nabla F(u), u - v > + |\lambda| \|u - v\|^2$$

Lemma 2.11 *Suppose that F_e satisfies $\mathcal{C}\wr\backslash\sqsubseteq$, then there exists an interval $I = (0, \hat{\lambda})$ such that in $c \in I$*

$$< \nabla F_e(u) - \nabla F_e(v), u - v > \geq c\|u - v\|^2, \forall u, v \in \mathbb{R}^n$$

Lemma 2.12 *Let $\hat{\lambda} = \sup()$ for all $u, v \in \mathbb{R}^n$, then*

$$\hat{\lambda} \geq |\lambda|$$

We can now state the stability result

Theorem 2.13 *Consider the RSS-convex splitting scheme*

$$\frac{u^{(k+1)} - u^{(k)}}{\Delta t} + \tau B(u^{(k+1)} - u^{(k)}) + \nabla F_c(u^{(k+1)}) = -Au^{(k)} + \nabla F_e(u^{(k)}), \tag{22}$$

Then

- *If* $(\tau\beta - \frac{1}{2}\rho(A) + \left(\hat{\lambda} - |\lambda|\right) > 0$ *then the scheme is unconditionally stable*
- *Else it is stable under condition*

$$0 < \Delta t < \frac{1}{(\frac{1}{2} - \tau\beta)\rho(A) + |\lambda| - \hat{\lambda}}$$

Proof. We have from Lemma 2.2.2, taking $u = u^{(k+1)}$ and $v = u^{(k)}$

$$F(u^{(k+1)}) - F(u^{(k)}) \leq < \nabla F(u^{(k+1}), u^{(k+1)} - u^{(k)} > + |\lambda| \|u^{(k+1)} - u^{(k)}\|^2$$

Hence

$$\begin{aligned} F(u^{(k+1)}) - F(u^{(k)}) \quad & \leq < \nabla F_c(u^{(k+1}) - \nabla F_e(u^{(k+1}), u^{(k+1)} - u^{(k)} > + |\lambda| \|u^{(k+1)} - u^{(k)}\|^2 \\ & + < \nabla F_e(u^{(k)}) - Au^{(k)} - \frac{1}{\Delta t}(u^{(k+1)} - u^{(k)}) - \tau < B(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > \\ & \leq - < \nabla F_e(u^{(k+1)}) - \nabla F_e(u^{(k)}), u^{(k+1)} - u^{(k)} > + |\lambda| \|u^{(k+1)} - u^{(k)}\|^2 \\ & - < Au^{(k)}, u^{(k+1)} - u^{(k)} > -\tau < B(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > \end{aligned}$$

We now use lemma 2.2.2-2.2.2 and the parallelogram identity, we get

$$\begin{aligned} F(u^{(k+1)}) - F(u^{(k)}) \quad & \leq - \left(\hat{\lambda} - |\lambda| + \frac{1}{\Delta t}\right) \|u^{(k+1)} - u^{(k)}\|^2 \\ & + \frac{1}{2} < Au^{(k)}, u^{(k)} > - \frac{1}{2} < Au^{(k+1)}, u^{(k+1)} > -\tau < B(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > \\ & + \frac{1}{2} < A(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > \end{aligned}$$

Finally, setting $E(u) = F(u) + \frac{1}{2} < Au, u >$, we find

$$E(u^{(k+1)}) + R \leq E(u^{(k)})$$

with

$$\begin{aligned} R \quad & = \left(\hat{\lambda} - |\lambda| + \frac{1}{\Delta t}\right) \|u^{(k+1)} - u^{(k)}\|^2 + \tau < B(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > \\ & - \frac{1}{2} < A(u^{(k+1)} - u^{(k)}), u^{(k+1)} - u^{(k)} > \end{aligned}$$

so the scheme is stable as $R > 0$. Hence the result. ■ Further discussions and implementations of RRS schemes to solve phase fields equations such as Allen Cahn's or Cahn Hilliard's will be considererd elsewhere. However, many important models can be included in the above general presentation (including Navier Stokes equations that are considered in Section 5). Hence we think that it is useful to introduce and analyze these time marching schemes in this part of the article.

2.3 Richardson extrapolation

In [1] the authors have proposed to increase the accuracy of the stabilized scheme by smoothing the residual using a Richardson extrapolation process:
The solution of

$$\frac{du}{dt} = F(u),$$

by the forward Euler scheme defines the iterations

$$u^{k+1} = u^k + \Delta t F(u^k) = G_{\Delta t}(u^k).$$

The smoothed sequence is defined by

$$\begin{aligned} v_1 &= G_{\Delta t}(u^k), \\ v_{2,0} &= G_{\Delta t/2}(u^k), \\ v_{2,1} &= G_{\Delta t/2}(v_{2,0}), \\ u^{k+1} &= 2v_{2,1} - v_1. \end{aligned}$$

It is second order accurate in time. The accuracy of the stabilized scheme is increased by applying the extrapolation to the iteration operator

$$G_{\Delta t,\tau}(u^k) = u^k + \Delta t(Id + \tau \Delta t B)^{-1} F(u^k).$$

The improvement is studied analytically in [21], but numerical evidences point out the efficiency of the approach, see also the numerical results presented in Section 4.

Below the Extrapoled RSS scheme

Algorithm 1 : Extrapoled RSS Scheme

```
$u^{(0)}$ given

for $k = 0, 1, \cdots$ until convergence do
    Solve $(Id + \tau\frac{\Delta t}{2}B)v_1 = -\frac{\Delta t}{2}F(u^{(k})$
    Set $u_1 = u^{(n)} + v_1$
    Solve $(Id + \tau\frac{\Delta t}{2}B)v_2 = -\frac{\Delta t}{2}F(u_1)$
    Set $u_2 = u_1 + v_2$
    Solve $(Id + \tau\Delta t B)v_3 = -\Delta t F(u^{(k)})$
    Set $u_3 = u^{(n)} + v_3$
    Set $u^{(k+1)} = 2u_2 - u_3$
end for
```

3 Discretisation in space and preconditioning

3.1 Compact FD Scheme Discretization

A way to obtain a high level of accuracy with a finite difference scheme, that can be compared with the spectral one, is to implement finite difference compact schemes [17]. These schemes consist in approaching a linear operator (differentiation, interpolation) by a rational (instead of polynomial-like) finite differences scheme. We describre briefly here only their construction for the approximation of the first and the second derivative and we refer to [17] for more details. We first consider the schemes in space dimension one.

Let $U = (U_1, \cdots, U_n)^T$ denotes a vector whose the components are the approximations of a regular function u at (regularly spaced) grid points $x_i = ih$, $i = 1, \cdots, n$. We compute approximations of $V_i = \mathcal{L}(u)(x_i)$ as solution of a system

$$P.V = QU,$$

so the approximation matrix is formally $B = P^{-1}Q$. When $P = Id$, the scheme is explicit and we recover the framework of classical finite difference schemes; when $P \neq Id$, the scheme is implicit an dit is possible to reach high order accuracy, the implicity allows to mimmic the spectral global dependence. In practice, P is a banded sparse matrix easy to invert and very well conditioned making the compact scheme approach robust and not costly regardless to the precision brought. We here give the matrices P and Q for the fourth order approximation of the first and the second derivative, details can be found in [17].

Let us begin with the first derivative. We have

$$P = \begin{pmatrix} 1 & \frac{1}{4} & & \\ \frac{1}{4} & 1 & \ddots & \\ & \ddots & \ddots & \frac{1}{4} \\ & & \frac{1}{4} & 1 \end{pmatrix},$$

$$Q = \frac{1}{2h}\begin{pmatrix} a_1 & a_2 & a_3 & a_4 & \\ -\frac{3}{2} & 0 & \frac{3}{2} & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{3}{2} & 0 & \frac{3}{2} \\ & -a_4 & -a_3 & -a_2 & -a_1 \end{pmatrix},$$

with $a_1 = -2$, $a_2 = 3$, $a_3 = -\frac{2}{3}$ and $a_4 = \frac{1}{8}$.

In the same way, we can build the fourth order compact schemes for the second order derivative. We first consider the compact scheme asociated to the discretization of the second derivative with homogeneous Dirichlet boundary conditions. We have with

$$P = \begin{pmatrix} 1 & \frac{1}{10} & & & \\ \frac{1}{10} & 1 & \frac{1}{10} & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{1}{10} & 1 & \frac{1}{10} \\ & & & \frac{1}{10} & 1 \end{pmatrix},$$

and

$$Q = \frac{1}{h^2}\begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & & \\ -\frac{6}{5} & \frac{12}{5} & -\frac{6}{5} & & & & \\ & -\frac{6}{5} & \frac{12}{5} & -\frac{6}{5} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -\frac{6}{5} & \frac{12}{5} & -\frac{6}{5} & \\ & & & & -\frac{6}{5} & \frac{12}{5} & -\frac{6}{5} \\ & & a_{N-4} & a_{N-3} & a_{N-2} & a_{N-1} & a_N \end{pmatrix},$$

here the constant a_1, a_2, a_3, ... are given by

$$\begin{cases} a_1 & = & a_N & = & -\frac{67}{60}, \\ a_2 & = & a_{N-1} & = & -\frac{7}{12}, \\ a_3 & = & a_{N-2} & = & \frac{13}{10}, \\ a_4 & = & a_{N-3} & = & -\frac{61}{120}, \\ a_5 & = & a_{N-4} & = & \frac{1}{12}. \end{cases}$$

Now applying the same approach, we can consider fourth order compact schemes for the second derivative with associated homogeneous Neumann Boundary conditions

$$M = \begin{pmatrix} 1 & \frac{1}{10} & & & \\ \frac{1}{10} & 1 & \frac{1}{10} & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{1}{10} & 1 & \frac{1}{10} \\ & & & \frac{1}{10} & 1 \end{pmatrix},$$

and

$$N = \frac{1}{h^2}\begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & & \\ -\frac{6}{5} & \frac{12}{5} & -\frac{6}{5} & & & & \\ & -\frac{6}{5} & \frac{12}{5} & -\frac{6}{5} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -\frac{6}{5} & \frac{12}{5} & -\frac{6}{5} & \\ & & & & -\frac{6}{5} & \frac{12}{5} & -\frac{6}{5} \\ & & a_{N-4} & a_{N-3} & a_{N-2} & a_{N-1} & a_N \end{pmatrix},$$

with

$$\begin{cases} a_1 &= & a_N &= & \frac{2681}{480}, \\ a_2 &= & a_{N-1} &= & -\frac{32}{3}, \\ a_3 &= & a_{N-2} &= & \frac{113}{40}, \\ a_4 &= & a_{N-3} &= & -\frac{13}{15}, \\ a_5 &= & a_{N-4} &= & \frac{59}{480}. \end{cases}$$

To obtain the compact schemes of first and second order derivative in space dimension 2 and 3, it suffices to use the previous schemes and to expand them tensorally.
The finite differences discretization matrices of derivative on cartesian domains can be obtained by those on the interval using Kronecker products. Indeed, if A^N_{xx} denotes the discretization matrix on $[0,1]$ associated to Dirichlet Boundary conditions, using N internal discretization points, then

$$Id_M \otimes A^N_{xx}$$

will be the discretization matrix of the same operator but on a grid composed of $N \times M$ points, the corresponding laplacian matrix will be $A^M_{xx} \otimes Id_N + Id_M \otimes A^N_{xx}$. We recall that the Kronecker product of a $m \times n$ matrix A by a $p \times q$ matix B is defined as

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}.$$

3.2 Preconditioning FD Compact schemes using second order FD

The matrices associated to compact finite difference schemes are full, this is due to the implicit nature of the scheme. A natural idea to built a sparse preconditioner is to use the matrix obtained by applying a lower accurate finite discretization scheme, particularly a second order one. We here describe the approach for one dimensional problem, extensions to higher dimensions are obtained using kronecker products, also we restrict to linear problem for simplicity. Let A_2 (resp A_4) be the second order (resp. the fourth order) discretization matrice of $-\Delta$ on a regular grid composed of N internal points. The RSS scheme writes as

$$\frac{u^{(k+1)} - u^{(k)}}{\Delta t} + \tau A_2(u^{(k+1)} - u^{(k)}) + A_4 u^{(k)} = f. \tag{23}$$

The numerical treatment of non homogeneous (possibly time depending) Dirichlet boundary conditions can be realized with the RSS approach. Indeed, let us note $A_m(u,n)$, $m = 2,4$, the mth order finite difference discretization of $-\Delta$ of u with Dirichlet conditions at time $n\Delta t$, note that this operator is affine. The stabilized scheme writes formally as

$$\frac{u^{(k+1)} - u^{(k)}}{\Delta t} + \tau(A_2(u^{(k+1)}, k+1) - A_2(u^{(k)}, k)) + A_4(u^{(k)}, k) = f, \tag{24}$$

Making the approximation $A_2(u^{(k+1)}, k+1) \simeq A_2(u^{(k+1)}, k)$, we obtain

$$\frac{u^{(k+1)} - u^{(k)}}{\Delta t} + \tau A_2(u^{(k+1)} - u^{(k)}) + A_4(u^{(k)}, k) = f. \tag{25}$$

It is to be pointed out that for the solutions of 2D and 3D Poisson problems, the number of iterations to the convergence is not dependent on the dimension of the system. Also, in the case of Poisson-like problem, we can use FFT to solve the preconditioning system, speeding up the resolution of the original linear system. This approach will be followed also for the fast solution of the heat equation that arises in parabolic problems.

Remark 3.1 *A analogous approach have been done in the context of hierarchical preconditioners in finite differences, where the fourth order discretization matrix of $-\Delta$ was preconditioned by the hierarchical transfert matrix attached to the second order accurate discretization of $-\Delta$, see [5].*

Below, we report numerical results on the solution of 2D and 3D Poisson problems when discretized by fourth order compact schemes. The preconditioning matrix is obtained by applying corresponding second order finite difference schemes. We took a random r.h.s `b=1-2*rand(N,1)` so that many frequencies including high ones are present. The initial data is $u = 0$, the tolerance parameter 10^{-12}. The result we report is the maximum number of external iterations to convergence, on 5 independent numerical resolutions, the number of discretization point per direction n is reported as (n). The stiffnes matrices are then of respective sizes $n^2 \times n^2$ (2D problem) and $n^3 \times n^3$ (3D problem).

Problem	# it. (n)	# it. (n))	# it. (n)	# it. (n)	#it. (n)	#it. (n)
Poisson 2D	12 (n=15)	11 (n=31)	10 (n=63)	10 (n=127)	9 (n=255)	8 (n=511)
Poisson 3D	12 (n=15)	11 (n=31)	11 (n=63)			

Table 1: Solutions of 2D and 3D Poisson problem with GMRES and second order preconditioner

Here, the fourth order discretization matrix A of $-\Delta$ is nonsymmetric while the preconditioning matrix B is. However, in practice the RSS method is still efficient. This is due to the small size of the skewsymmetric part of A. Indeed, denoting by $\delta =\| A - A^T \|$, we can prove the following general stability result.

Theorem 3.2 *Let $A \in \mathcal{M}_n(\mathbb{R}^N)$. We assume that A is positive definite and B a symmetric definite positive preconditioning matrix of A satisfy hypothesis $\mathcal{H}$. We set $\delta =\| A - A^T \|$ and $\Phi(\xi) = (\beta^2 - 2\alpha\tau)\xi + \frac{1}{4\xi}\delta^2$. Assume that $\frac{\beta^2}{2\alpha} - \frac{\delta^2}{8\alpha\lambda_{min}(B)^2} \geq 0$. Then the RSS scheme has the following stability conditions*

i. if $\tau \geq \frac{\beta^2}{2\alpha} + \frac{\delta^2}{8\alpha\lambda_{min}^2(B)} \geq \frac{\beta^2}{2\alpha}$. then the scheme is unconditionally stable.

ii. If $\tau \leq \frac{\beta^2}{2\alpha} - \frac{\delta^2}{8\alpha\lambda_{max}(B)^2}$ then the scheme is table under condition

$$0 < \Delta t < \frac{2\alpha}{\Phi(\lambda_{max}(B))}$$

iii. If $\frac{\beta^2}{2\alpha} - \frac{\delta^2}{8\alpha\lambda_{max}(B)^2} \leq \tau < \frac{\beta^2}{2\alpha} + \frac{\delta^2}{8\alpha\lambda_{min}(B)^2}$ *then the scheme is table under condition*

$$0 < \Delta t < \frac{2\alpha}{\Phi(\lambda_{min}(B))}$$

iv. If $\frac{\beta^2}{2\alpha} - \frac{\delta^2}{8\alpha\lambda_{min}(B)^2} < \tau < \frac{\beta^2}{2\alpha} - \frac{\delta^2}{8\alpha\lambda_{max}(B)^2}$ *then the scheme is table under condition*

$$0 < \Delta t < \frac{2\alpha}{\text{Max}(\Phi(\lambda_{min}(B)), \Phi(\lambda_{max}(B)))}$$

Here $\lambda_{min}(B)$ (resp. $\lambda_{max}(B)$ denotes the lowest (resp. the largest) eigenvalue of B.

Proof. The RSS scheme reads as

$$u^{(k+1)} = u^{(k)} - \Delta t \left(Id + \tau\delta t B\right)^{-1} A u^{(k)} = M u^{(k)},$$

and is stable under the necessary and sufficient condition $\rho(M) < 1$; in the general case the eigenvalues of M can be complex. Let $v \in \mathbb{C}^N$ be an eigenvector of M associated to the eigenvalue $\lambda = a + ib$. We have

$$\lambda v = Mv, \text{ so } (1-\lambda)\left(Id + \tau\delta t B\right) v = \Delta t A v.$$

We decompose v as $v = v_1 + iv_2$ and, separating real and imaginary parts, we obtain

$$\begin{aligned} Av_1 &= \frac{1-a}{\Delta t}\left(v_1 + \tau\Delta t B v_1\right) + \frac{b}{\Delta t}\left(v_2 + \tau\Delta t B v_2\right), \\ Av_1 &= \frac{1-a}{\Delta t}\left(v_2 + \tau\Delta t B v_2\right) - \frac{b}{\Delta t}\left(v_1 + \tau\Delta t B v_1\right). \end{aligned}$$

We have then, on the one hand

$$< Av_1, v_1 > + < Av_2, v_2 >= \frac{1-a}{\Delta t}\left(\| v_1 \|^2 + \| v_2 \|^2 + \tau\Delta t < Bv_1, v_1 > + \tau\Delta t < Bv_2, v_2 >\right),$$

and, on the other hand,

$$< Av_1, v_2 > - < Av_2, v_1 >= \frac{b}{\Delta t}\left(\| v_1 \|^2 + \| v_2 \|^2 + \tau\Delta t < Bv_1, v_1 > + \tau\Delta t < Bv_2, v_2 >\right).$$

We now set for convenience $N(v_1, v2) =\| v_1 \|^2 + \| v_2 \|^2 + \tau\Delta t < Bv_1, v_1 > + \tau\Delta t < Bv_2, v_2 >$. We infer from the previous identities

$$a = 1 - \frac{\Delta t}{N(v_1, v_2)}\left(< Av_1, v_1 > + < Av_2, v_2 >\right) \text{ and } b = \frac{\Delta t}{N(v_1, v_2)} < (A - A^T)v_1, v_2 > .$$

The stability condition is

$$\eta = a^2 + b^2 < 1. \tag{26}$$

After the usual simplifications, we find as a sufficient condition

$$\begin{aligned}\eta \quad & \leq 1 - 2\frac{\Delta t}{N(v_1,v_2)}\alpha\left(< Bv_1, v_1 > + < Bv_2, v_2 >\right) \\ & + \frac{\Delta t^2}{N(v_1,v_2)^2}\beta^2\left(< Bv_1, v_1 > + < Bv_2, v_2 >\right)^2 \\ & + \frac{\Delta t^2}{4N(v_1,v_2)^2} \parallel A - A^T \parallel^2 \left(\parallel v_1 \parallel^2 + \parallel v_2 \parallel^2\right)^2 .\end{aligned}$$

We now let $Z =< Bv_1, v_1 > + < Bv_2, v_2 >$ and $Y =\parallel v_1 \parallel^2 + \parallel v_2 \parallel^2$. The last inequality reads

$$\Delta t \left((\beta^2 - 2\alpha\tau)Z^2 + \frac{1}{4} \parallel A - A^T \parallel^2 Y^2 \right) < 2\alpha Y Z.$$

At this point, we set $\xi = \frac{Z}{Y} = \frac{< Bv_1, v_1 > + < Bv_2, v_2 >}{\parallel v_1 \parallel^2 + \parallel v_2 \parallel^2}$. Note that we have $\lambda_{min}(B) \leq \xi \leq \lambda_{max}(B)$. After usual simplifications, the sufficient stability condition (26) writes now as

$$\Delta t \left((\beta^2 - 2\alpha\tau)\xi + \frac{1}{4}\delta^2\frac{1}{\xi} \right) < 2\alpha,$$

where we have set $\delta =\parallel A - A^T \parallel$. We use the function $\Phi(\xi) = (\beta^2 - 2\alpha\tau)\xi + \frac{1}{4\xi}\delta^2$; Φ is obviously regular on $[\lambda_{min}(B), \lambda_{max}(B)]$, since $\lambda_{min}(B) > 0$, B is assumed to be SPD. We deduce the following sufficient stability conditions

- if $\Phi(\xi) \leq 0, \forall \xi \in [\lambda_{min}(B), \lambda_{max}(B)]$, the scheme is unconditionnaly stable
- if there exists $\xi \in [\lambda_{min}(B), \lambda_{max}(B)]$ such that $\Phi(\xi) > 0$, then a sufficient stablility condition is
$$0 < \Delta t < \frac{2\alpha}{\underset{\xi\in[\lambda_{min}(B),\lambda_{max}(B)]}{\text{Max}} \Phi(\xi)}.$$

We conclude by studying the two cases.

Unconditional stablity : To have $\Phi(\xi) \leq 0, \forall \xi \in [\lambda_{min}(B), \lambda_{max}(B)]$, we must have $\Phi(\lambda_{min}(B)) \leq 0$ and $\Phi(\lambda_{max}(B)) \leq 0$, that is

$$\tau \geq \frac{\beta^2}{2\alpha} + \frac{\delta^2}{8\alpha\lambda_{min}^2(B)} \geq \frac{\beta^2}{2\alpha}.$$

We now remark that $\Phi'(\xi) = (\beta^2 - 2\alpha\tau) - \frac{\delta^2}{4\xi^2} \leq 0 \iff \tau \geq \frac{\beta^2}{2\alpha} - \frac{\delta^2}{8\alpha\xi^2}, \forall \xi \in [\lambda_{min}(B), \lambda_{max}(B)]$. This is satisfied under the previous hypothesis, hence the first statement $[i.]$ of the theorem.

Conditional stability : The condition is $\text{Max}(\Phi(x)) > 0$. We distinguish two cases

- Φ is monotone.

  - $\Phi'(\xi) \geq 0$ and $\Phi(\lambda_{max}(B)) > 0$ ie $\tau < \frac{\beta^2}{2\alpha} + \frac{\delta^2}{8\alpha\lambda_{max}(B)^2}$ and $\tau \leq \frac{\beta^2}{2\alpha} - \frac{\delta^2}{8\alpha\lambda_{max}(B)^2}$.

    A sufficent stability condition is then $\tau \leq \frac{\beta^2}{2\alpha} - \frac{\delta^2}{8\alpha\lambda_{max}(B)^2}$ and
    $$0 < \Delta t < \frac{2\alpha}{\Phi(\lambda_{max})},$$

  - $\Phi'(\xi) \leq 0$ and $\Phi(\lambda_{min}(B)) > 0$ that is $\frac{\beta^2}{2\alpha} - \frac{\delta^2}{8\alpha\lambda_{max}(B)^2} \leq \tau < \frac{\beta^2}{2\alpha} + \frac{\delta^2}{8\alpha\lambda_{min}(B)^2}$ and
    $$0 < \Delta t < \frac{2\alpha}{\Phi(\lambda_{min}(B))},$$

- $\Phi'(\xi) = 0$ for $\xi \in]\lambda_{min}(B), \lambda_{max}(B)[$. This means that
  $$\frac{\beta^2}{2\alpha} - \frac{\delta^2}{8\alpha\lambda_{min}(B)^2} < \tau < \frac{\beta^2}{2\alpha} - \frac{\delta^2}{8\alpha\lambda_{max}(B)^2}.$$
  This implies that $\Phi'(\lambda_{min}(B)) < 0$ so $\text{Max}(\Phi(\xi))$ is reached at $\lambda_{min}(B)$ or at $\lambda_{max}(B)$ and the stability condition is
  $$0 < \Delta t < \frac{2\alpha}{\text{Max}(\Phi(\lambda_{min}(B)), \Phi(\lambda_{max}(B)))}.$$

■ We point out that the symmetric part of matrix A is dominant for small values of δ and that in this case the above stability result is to be compared with that of Proposition (2.2.1). Particularly inconditional stability condition is obtained for $\tau \geq \frac{\beta^2}{2\alpha} + \frac{\delta^2}{8\alpha\lambda_{min}^2(B)} \simeq \frac{\beta}{2}$, for inconditional preconditioners B as those presented above.

4 Numerical Results

4.1 The problem and the numerical schemes

As stated in the introduction, we here aim at computing the steady state of the so-called 2D driven cavity problem in a rectangular domain Ω. The steady state will be reached by a pseudo-temporal method, and for that purpose we consider the stream function-vorticity formulation $(\omega - \psi)$ (see [19, 24] and the references therein):

$$\frac{\partial \omega}{\partial t} - \frac{1}{Re}\Delta\omega + \frac{\partial \phi}{\partial y}\frac{\partial \omega}{\partial x} - \frac{\partial \phi}{\partial x}\frac{\partial \omega}{\partial y} = 0, \quad \text{in } \Omega \tag{27}$$

$$\Delta\psi = \omega, \quad \text{in } \Omega \tag{28}$$

$$\omega(x, y, 0) = \omega_0(x, y), \tag{29}$$

that we supplement with proper boundary conditions. We denote by $\Gamma_i \;\; i = 1, .., 4$ the sides of the unit square Ω as follows: Γ_1 is the lower horizontal side, Γ_3 is the upper horizontal side, Γ_2 is the left vertical side, and Γ_4 is the right vertical side.

We distinguish two different driven flows, according to the choice of the boundary conditions on the velocity. More precisely we have

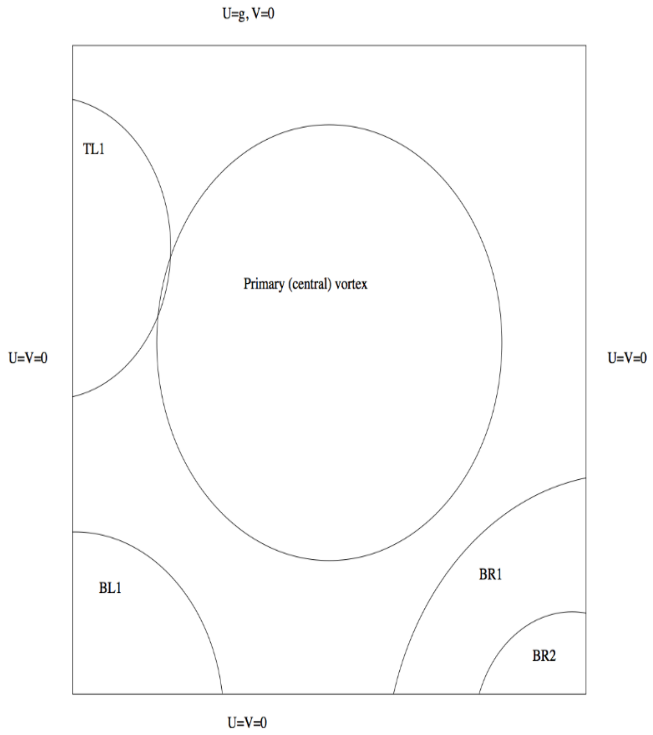

Figure 2: The lid driven cavity - Schematic localization of the mean vortex regions

- $g(x) = 1$: Cavity A (lid driven cavity)
- $g(x) = (1 - (1 - 2x)^2)^2$: Cavity B (regularized lid driven cavity)

We will also consider the steady linear part of this equation (the Stokes problem), whose the solution will be chosen as the initial condition of (27)

$$-\frac{1}{Re}\Delta\omega = 0, \quad \text{in } \Omega \tag{30}$$

$$\omega \mid_{\partial\Omega} = \Delta\psi \mid_{\partial\Omega} \quad \text{on}\partial\Omega, \tag{31}$$

$$\Delta\psi = \omega, \quad \text{in } \Omega \tag{32}$$

$$\psi \mid_{\partial\Omega} = 0 \quad \text{on } \partial\Omega, \tag{33}$$

The RSS scheme is based on two different finite differences discretization of differential operators at the same grid points: a second order finite difference scheme will be used for preconditiong while the fourth order compact scheme is implemented for the effective approximation to the solution.

The boundary conditions on ω are derived by the discretization of $\Delta\psi$ on the boundaries.

With the conditions on u and v we have

$$\begin{aligned}
\omega(x,0,t) &= \frac{\partial^2\psi}{\partial y^2}(x,0,t) \quad \text{on } \Gamma_1,\\
\omega(x,1,t) &= \frac{\partial^2\psi}{\partial y^2}(x,1,t) \quad \text{on } \Gamma_3,\\
\omega(0,y,t) &= \frac{\partial^2\psi}{\partial x^2}(0,y,t) \quad \text{on } \Gamma_2,\\
\omega(1,y,t) &= \frac{\partial^2\psi}{\partial x^2}(1,y,t) \quad \text{on } \Gamma_4.
\end{aligned}$$

So, since $\psi_{\partial\Omega} = 0$ and $u = \dfrac{\partial\psi}{\partial y}$, $v = -\dfrac{\partial\psi}{\partial x}$, we obtain by using Taylor expansions

$$\begin{aligned}
\omega_{i,0} &= \frac{\psi_{i,1} - 8\psi_{i,2}}{2h^2},\\
\omega_{i,N+1} &= \frac{-\psi_{i,N-1} + 8\psi_{i,N} - 6hg(ih)}{2h^2},\\
\omega_{0,j} &= \frac{\psi_{1,j} - 8\psi_{2,j}}{2h^2}i,\\
\omega_{N+1,j} &= \frac{-\psi_{N-1,j} + 8\psi_{N,j}}{2h^2}.
\end{aligned} \tag{34}$$

Here $g(x)$ denotes the boundary condition function for the horizontal velocity at the boundary Γ_3.

The boundary conditions on ψ are homogeneous Dirichlet BC. The operators are discretized by second order centered schemes on a uniform mesh composed by N points in each direction of the domain of step-size $h = \dfrac{1}{N+1}$. The total number of unknowns is then $2N^2$.

The boundary conditions on ω are iteratively implemented according to the relations (34-34), making the finite differences scheme second order accurate. Using the following fourth order accurate extrapolation,

$$\left\{\begin{aligned}
\omega_{i,0} &= \frac{1}{h^2}\left(8\psi_{i,1} - 3\psi_{i,2} + \frac{8}{9}\psi_{i,3} - \frac{1}{8}\psi_{i,4}\right),\\
\omega_{i,N+1} &= \frac{1}{h^2}\left(8\psi_{i,N} - 3\psi_{i,N-1} + \frac{8}{9}\psi_{i,N-2} - \frac{1}{8}\psi_{i,N-3}\right) - \frac{25}{6h}g(ih),\\
\omega_{0,j} &= \frac{1}{h^2}\left(8\psi_{1,j} - 3\psi_{2,j} + \frac{8}{9}\psi_{3,j} - \frac{1}{8}\psi_{4,j}\right),\\
\omega_{N+1,j} &= \frac{1}{h^2}\left(8\psi_{N,j} - 3\psi_{N-1,j} + \frac{8}{9}\psi_{N-2,j} - \frac{1}{8}\psi_{N-3,j}\right),
\end{aligned}\right. \tag{35}$$

we complete the discretization; one can refer also to [10, 18]. Now the implementation of the RSS scheme reads as

- Convection-Diffusion problem: knowing $\psi^{(k)}$, compute $\omega^{(k+1)}$ solution of

$$\left\{\begin{aligned}
\frac{\omega^{(k+1)} - \omega^{(k)}}{\Delta t} - \frac{1}{Re}\Delta\omega^{(k+1)} + \frac{\partial\psi^{k}}{\partial y}\frac{\partial\omega^{(k)}}{\partial x} - \frac{\partial\psi^{(k)}}{\partial x}\frac{\partial\omega^{(k)}}{\partial y} &= 0 \qquad \text{in } \Omega =]0,1[^2\\
\omega^{(k+1)} &= \Delta\psi^{(k)} \quad \text{on } \partial\Omega
\end{aligned}\right. \tag{36}$$

- Poisson problem: knowing $\omega^{(k+1)}$, compute $\psi^{(k+1)}$ solution of

$$\left\{ \begin{array}{lcll} \omega^{(k+1)} & = & \Delta\psi^{(k+1)} & \text{in } \Omega =]0,1[^2, \\ \psi^{(k+1)} & = & 0 & \text{on}\partial\Omega. \end{array} \right. \tag{37}$$

Algorithm 2 : RRS Navier-Stokes

```
$(\omega^0, \psi^0)$ given as solution of the Stokes problem (30)
**for** $k = 0, 1, \cdots$ until convergence **do**
    Update the boundary terms using (35)
    Compute $\omega^{(k+1)}$ by solving. (36) with RSS (7)
    Compute $\psi^{n+1}$as solution of the Poisson equation (37)
**end for**
```

4.2 RSS Schemes for computing Steady States of the lid driven cavity

We give now numerical results on the square cavity $\Omega =]0,1[^2$, we compare the numerical values of the steady state with those of the literature. Here $g(x) = 1$. There is an enormous amount of works dealing with the numerical simulation of the lid driven cavity which still is a benchmark. We chose to compare qualitatively our results with those presented in [3, 4, 15, 16], however we can find analysis of compact schemes and numerical simulatiosn of the unsteady and steady lid driven cavity , e.g., in [10, 18] and in the references there in.

The steady state is computed for $\|\frac{\partial\psi}{\partial t}\| < \varepsilon = 10^{-3}$. We report hereafter the vorticity and the stream function in figures 3, 4, 5 and 6, for $Re = 100, Re = 400, Re = 1000, Re = 3200$ respectively. They agree with those of the literature [3, 4, 15, 16]; the localization of the extrema of ω and ψ are reported on table 2.

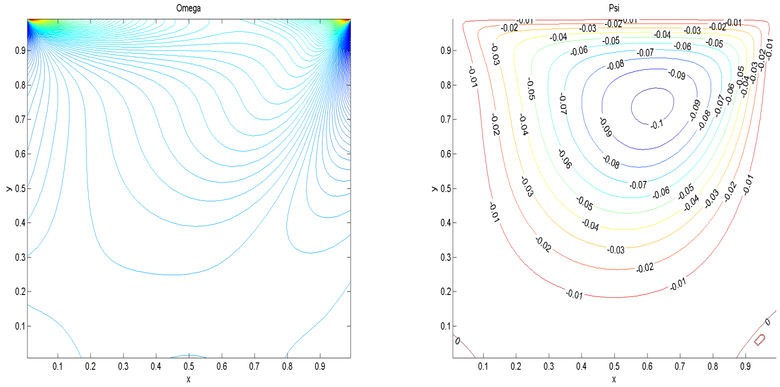

Figure 3: Steady solution of NSE (27) - $g \equiv 1$ - $\tau = 1$ - $N = 127$ - $Re = 100$ - $\Delta t = 0.001$

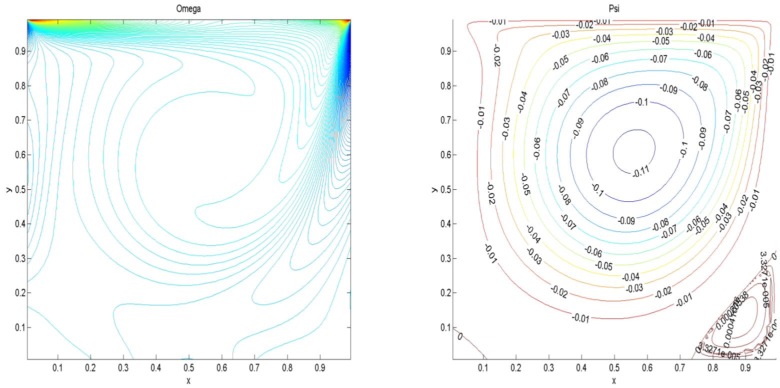

Figure 4: Solution of NSE (27) - $g \equiv 1$ - $\tau = 1$ - $N = 127$ - $Re = 400$ - $\Delta t = 0.001$

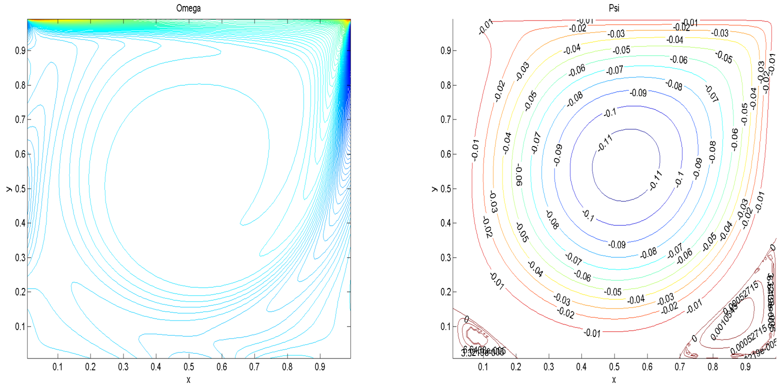

Figure 5: Solution of NSE (27) - $g \equiv 1$ - $\tau = 1$ - $N = 127$ - $Re = 1000$ - $\Delta t = 0.0005$

Now we illustrate the influence of the stabilization parameter τ on the convergence in time to the steady state. A large value of τ allows to take a large time step Δt but slows down the convergence in time. We have plotted in figures (7) and (8) the evolution in time of $\|\frac{\partial \psi}{\partial t}\|$. We observe that, for $\tau = 1$ the RSS schemes (first and second order) behave similarly as the reference one (semi backward Euler's); for $\tau = 100$, we see that RSS is slow downed while its extrapolated version has comparable dynamics to Euler's.

We now give numerical results for the rectangular cavity. They are presented in figures (9) ,(10), (11) and (12) for $Re = 100$, $Re = 400$, $Re = 1000$, $Re = 3200$ respectiveley. They agree with those obtained by Goyon [16], see also the numerical values reported in Table 3.

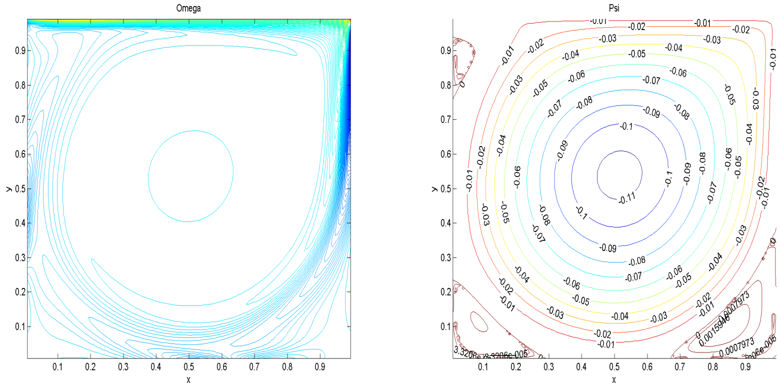

Figure 6: Solution of NSE (27) - $g \equiv 1$ - $\tau = 1$ - $N = 127$ - $Re = 3200$ - $\Delta t = 0.0005$

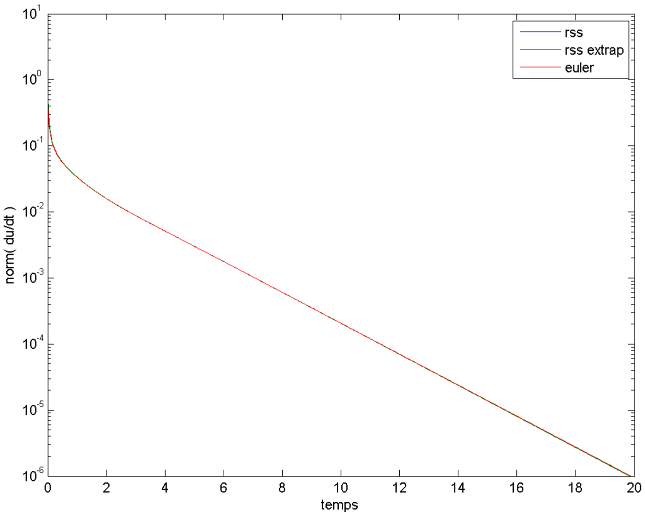

Figure 7: Convergence to NSE steady state (27) - $\tau = 1$ - $N = 63$ - $Re = 100$ - $\Delta t = 0.01$

$Re = 100$

	Principal Vortex	Ben Artzi *et al* [3]	O. Goyon [16]	extrapoled RSS
Spatial accuracy		4th order compact scheme	second order	4th order
grid		97×97	129×129	127×127
Δt			0.005	0.004
	intensity		0.1033	0.1026
	x		0.6172	0.6172
	y		0.7343	0.7422

$Re = 400$

	Principal Vortex	Ben Artzi *et al* [3]	O. Goyon [16]	extrapoled RSS
Spatial accuracy		4th order compact scheme	second order	4th order
grid		97×97	129×129	127×127
Δt				0.017
	intensity	0.1136		0.1123
	x	0.5521		0.5625
	y	0.6042		0.6094

$Re = 1000$

	Principal Vortex	Ben Artzi *et al* [3]	O. Goyon [16]	extrapoled RSS
Spatial accuracy		4th order compact scheme	second order	4th order
grid		97×97	129×129	127×127
Δt			0.02	0.01
	intensity	0.1178	0.1157	0.1158
	x	0.5312	0.5312	0.5391
	y	0.5625	0.5625	0.5703

$Re = 3200$

	Principal Vortex	Ben Artzi *et al* [3]	O. Goyon [16]	extrapoled RSS
Spatial accuracy		4th order compact scheme	second order	4th order
grid		97×97	129×129	127×127
Δt			0.01	0.006
	intensity	0.1174	0.1122	0.1129
	x	0.5208	0.5234	0.5156
	y	0.5417	0.5468	0.5391

Table 2: Extremal values of ψ and their localization for NSE (27). Comparison with results of Croisille and Goyon, for different Reynolds numbers. $g \equiv 1$ - $\tau = 1$

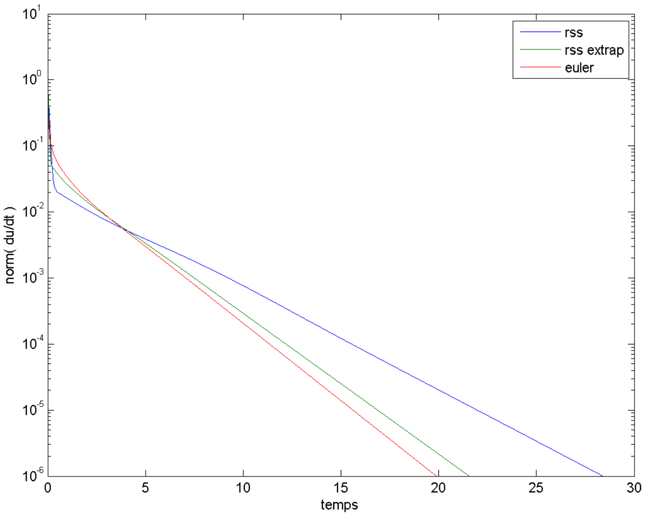

Figure 8: Convergence to NSE steady state (27) - $\tau = 100$ - $N = 63$ - $Re = 100$ - $\Delta t = 0.01$

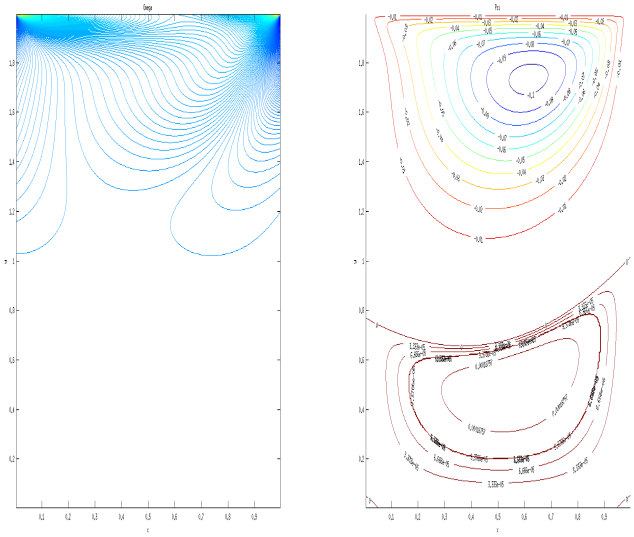

Figure 9: Solution of (27) in $[0; 1] \times [0; 2]$ - $g \equiv 1$ - $\tau = 1$ - 255×511 - $Re = 100$ - $\Delta t = 0.001$

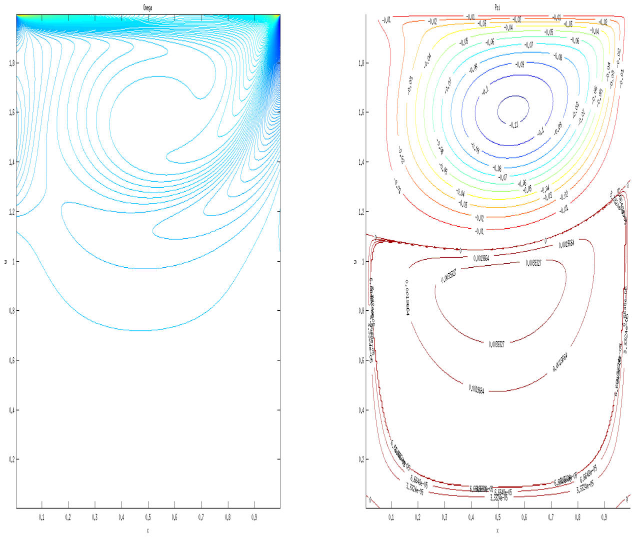

Figure 10: Solution of NSE (27) in $[0;1] \times [0;2]$ - $g \equiv 1$ - $\tau = 1$ - 255×511 - $Re = 400$ - $\Delta t = 0.001$

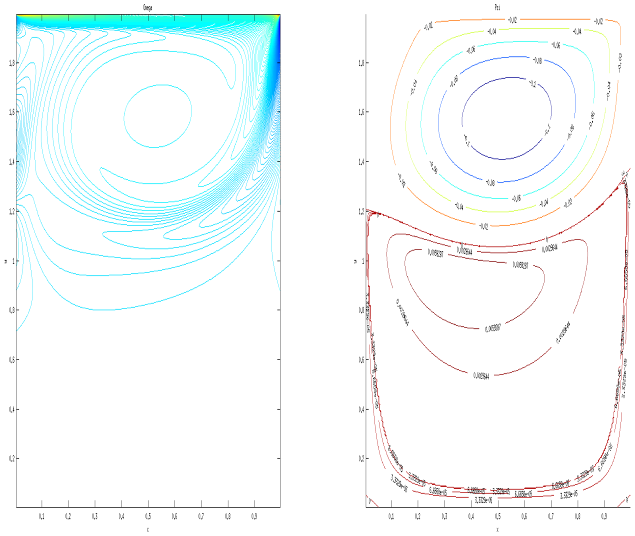

Figure 11: Solution of NSE (27) in $[0;1] \times [0;2]$ - $g \equiv 1$ - $\tau = 1$ - 255×511 - $Re = 1000$ - $\Delta t = 0.0005$

$Re = 100$

		extrapoled RSS	extrapoled RSS	O. Goyon [16]	C-H. Bruneau and
				second order	C. Jouron [4]
grid:		63×127	255×511	65×129	multi-grid
Δt		1.10^{-3}	1.10^{-3}	1.10^{-2}	
ϵ:		10^{-5}	10^{-3}	10^{-5}	
VS	ψ	0.1040	0.1034	0.1035	0.1033
	x	0.6094	0.6172	0.6093	0.6172
	y	1.7344	1.7344	1.7343	1.7344
VI	ψ	8×10^{-3}	0.0003	6.65×10^{-4}	0.783×10^{-3}
	x	0.5469	0.5820	0.5468	0.5391
	y	0.5938	0.5039	0.5781	0.5859

$Re = 400$

		extrapoled RSS	extrapoled RSS	O. Goyon [16]	C-H. Bruneau and
				second order	C. Jouron [4]
grid:		63×127	255×511	65×129	multi-grid
Δt		1.10^{-3}	1.10^{-3}	1.10^{-2}	
ϵ:		10^{-5}	10^{-3}	10^{-5}	
VS	ψ	0.1131	0.1120	0.1097	0.1124
	x	0.5469	0.5586	0.5625	0.5547
	y	1.6094	1.6094	1.6094	1.5938
VI	ψ	0.009	0.0059	8.06×10^{-3}	0.909×10^{-2}
	x	0.4219	0.5156	0.4375	0.4297
	y	0.8438	0.8750	0.8593	0.8125

$Re = 1000$

		extrapoled RSS	extrapoled RSS	O. Goyon [16]	C-H. Bruneau and
				second order	C. Jouron [4]
grid:		127×257	255×511	257×513	multi-grid
Δt		5.10^{-4}	5.10^{-4}	5.10^{-3}	
ϵ:		10^{-5}	10^{-5}	10^{-5}	
VS	ψ	0.1189	0.1196	0.1187	0.1169
	x	0.5312	0.5312	0.5313	0.5273
	y	1.5781	1.5781	1.5781	1.5625
VI	ψ	0.0134	0.0135	1.32×10^{-2}	0.0148
	x	0.3438	0.3398	0.3359	0.3516
	y	0.8438	0.8438	0.8476	0.7891

Table 3: Some extremal values of ω and ψ for the steady state of NSE (27) on $[0;1] \times [0;2]$ (upper and lower vortex) - $g \equiv 1$ - $\tau = 1$

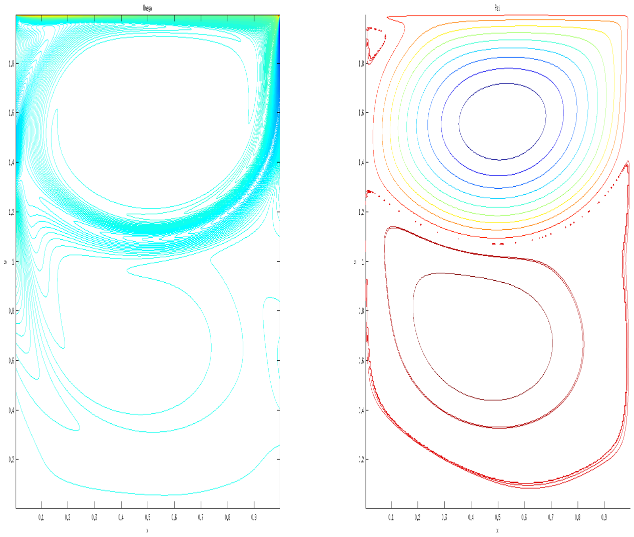

Figure 12: Solution of NSE (27) on $[0;1] \times [0;2]$ - $g \equiv 1$ - $\tau = 1$ - 255×511 - $Re = 3200$ - $\Delta t = 0.0005$

Here the vorticity ω and the streamfunction ψ at steady state for $Re = 3200$. The maximal values found for ψ are, for the maximum 0.0196 located in $(x, y) = (0.4492, 06914)$ and the minimum is -1.215, located in $(x, y) = (0.5156, 1.15664)$.

4.3 On the role of τ and of the extrapolation to the convergence in time to the steady state

We now denote by T_c the time at which the convergence to the steady state is reached, say $T_c = (k+1)\Delta t$ with $\| \dfrac{\psi^{(k+1)} - \psi^{(k)}}{\Delta t} \| \leq \epsilon$. We put the symbol $***$ when the scheme is unstable and blows up numerically; 'NC' when it is not convergent after A given large time (T=2000) and finally 'Inc. Stab' when it is inconditionally stable.

We first consider the RRS scheme with the second order laplacian matrix as a preconditioner. We show for low Reynolds numbers, that a large value of τ allows to use large Δt and that the slower convergence to the steady state can be corrected by using the Richardson extrapolation.

Of course, since the stabilization allows to take larger time steps, a important gain in CPU time can be obtained when computing a steady state. It can be estimated by considering the number of iteration in time NT: $\dfrac{T_c}{\Delta t}$ for RSS and $3\dfrac{T_c}{\Delta t}$ for RSS with extrapolation. For example, taking $Re = 400$ and $n = 127$, we find $NT = 2501$ for $\tau = 1$ and $\Delta t = 0.014$ and for $\tau = 30$ and $\Delta t = 0.6$, we have $NT = 410$ (RSS) and $NT = 568$ (RSS with extrapolation) ; hence a factor 6.1 is reached for RSS and one of 4.4 for extrapolated RSS, see Table 5.

τ	Δt	Δt_{max}	T_c	Δt	Δt_{max}	T_c
Extrapolation	no	no	no	yes	yes	yes
$\tau = 1$	0.01	0.019	15.62	0.01	0.019	15.6
	0.019		15.665	0.019		16.492
$\tau = 10$	0.02	0.14	16.86	0.02	0.11	15.8
	0.06		19.32	0.06		16.8
	0.07		19.95	0.07		17.29
	0.1		21.7	0.1		21.7
$\tau = 30$	0.3	Inc. Stab.	54.5	0.3	1.08	32.1
	0.4		79.6	0.4		38

Table 4: RSS (left and RSS with Extrapolation (right) $Re = 100$, $n = 63$, $\epsilon = 10^{-5}$

τ	Δt	Δt_{max}	T_c	Δt	Δt_{max}	T_c
Extrapolation	no	no	no	yes	yes	yes
$\tau = 1$	0.01	0.014	35.02	0.01	0.014	35.03
	0.014		35.014	0.014		35.042
$\tau = 20$	0.2	0.28	67.60	0.21	0.22	57.6
$\tau = 30$	0.3		105	0.3	0.35	58.2
	0.35		117.95	0.35		92.4
$\tau = 50$	0.3	5.1	139.2	0.3	1	69
	0.4		176.8	0.4		83.6
	0.5		212.5	0.5		99
	0.6		246.6	0.6		113.4

Table 5: RSS (left) and RSS witht Extrapolation (right) $Re = 400$, $n = 127$, $\epsilon = 10^{-5}$

We now consider larger Reynolds numbers and take into account the convective part of the equation in the construction of the sparse RSS preconditioner and apply nonlinear RSS scheme (6) to the vorticity time marching, say

$$\frac{\omega^{(k+1)} - \omega^{(k)}}{\Delta t} + \left(\frac{1}{Re}A_2 + diag(Dy\psi^{(k)})Dx - diag(Dx\psi^{(k)})Dy\right)(\omega^{(k+1)} - \omega^{(k)}) = -F(\psi^{(k)}, \omega^{(k)}) \tag{38}$$

where A_2 is the second order laplacian matrix, $diag(Dy\psi^{(k)})$ (resp. $diag(Dx\psi^{(k)})$) is the diagonal matrix with the discrete (second order accurate) approximation of $\frac{\partial \psi^{(k)}}{\partial x}$ (resp. $\frac{\partial \psi^{(k)}}{\partial y}$) at grid points as entries; Dx (resp. Dy) denote the (second order accurate) first derivative matrix in x (resp. in y) on the cartesian grid. Finally $-F(\psi^{(k)}, \omega^{(k)})$ is the high order compact scheme discretisation of $-\frac{1}{Re}\Delta\omega + \frac{\partial \phi}{\partial y}\frac{\partial \omega}{\partial x} - \frac{\partial \phi}{\partial x}\frac{\partial \omega}{\partial y}$.

As Shown on Table 6 and Table 7, NLRSS outperform RSS for the computation of Steady

Method	RSS	RSS	RSS	RSS	RSS	RSS	NLRSS	NLRSS	NLRSS
τ	Δt	Δt_{max}	T_c	Δt	Δt_{max}	T_c	Δt	Δt_{max}	T_c
Extrapolation	no	no	no	yes	yes	yes	yes	yes	yes
$\tau = 1$	0.005	0.005	56.21	0.005	0.01	56.81			
	0.01		***	0.01	56.79		0.01	0.02	56.86
	0.02		***	0.02	***		0.02		56.96
$\tau = 30$	0.05	0.04	NC	0.05	0.08	47.95	0.05	0.7	65.05
	0.1		***	0.1		***	0.1		62.5
	0.7		***	0.7		***	0.7		321.3

Table 6: RSS (left) RSS with Extrapolation (center) and extrapolated NLRSS (right) $Re = 1000$, $n = 127$, $\epsilon = 10^{-5}$

τ	Δt	Δt_{max}	T_c	Δt	Δt_{max}	T_c
Extrapolation	no	no	no	no	no	no
$\tau = 10$	0.1	0.6	223.9	0.1	0.006	***

Table 7: NLRSS with (left) and RSS (right) $Re = 3200$, $n = 127$, $\epsilon = 10^{-5}$

States for higher Reynolds numbers, for $Re = 1000$ and *a fortiori* for $Re = 3200$. It allows to use large times steps while RSS is unstable for such Δt.

5 Concluding remarks

We have studied RRS-like scheme (and their implementations) and pointed out their advantages for the numerical solution of parabolic problems when using high order compact schemes in finite differences for the space disctretization. In particular, the possibility of using fast solvers attached to a standard second order discretization, speeds up the resolution while bringing an enhanced stability. We also pointed out the role of the approximation of τB to A in the dynamics of the convergence to a steady state: a too strong stablization slows down the convergence in time while enhacing the stability of the scheme, the application of Richardson extrapolation allows to recover a dynamics close to the one of the classical schemes. The robustness of the schemes is illustrated with the solution of 2D NSE equations. The RSS approach is very versatile and allows adaptations of a large number of techniques of numerical analysis of ODEs. Many developments remain to consider, such as applying factorization updatings on the preconditioners or deriving and applying multilevel general (or Block) RSS schemes for the solution of other large scale parabolic problems. Applications of the RSS-approach to the solution of phase fields models is also a promising issue.

Acknoledgements: The authors thanks the anonymous referee for the careful reading and the constructive remarks that helped to improve the paper.

References

[1] A.Averbuch, A. Cohen, M.Israeli, A fast and accurate multiscale scheme for parabolic equations rapport LAN 1998.

[2] S. Bellavia, B. Morini, M. Porcelli, New Updates of incmplete LU factorizations and applications to large nonlinear systems, Optimization Methods & software, vol 29 pp 321-340, (2014).

[3] M. Ben-Artzi, J.–P. Croisille, D. Fishelov, S. Trachtenberg, A pure-compact scheme for the streamfunction formulation of NavierStokes equations, Journal of Computational Physics 205 (2005) 640664

[4] Ch.-H. Bruneau, C. Jouron, An efficient scheme for solving steady incompressible Navier-Stokes equations, J. of Comp. Physics Vol. 89, n 2, 1990.

[5] J.-P. Chehab, Incremental Unknowns Method and Compact Schemes , M2AN, 32, 1, (1998), 51-83.

[6] C. Calgaro, J.-P. Chehab, Y. Saad, Incremental Incomplete LU factorizations with applications to PDES, Numerical Linear Algebra with Applications, vol 17, 5, p 811–837, 2010.

[7] J.-P. Chehab et B. Costa, Multiparameter schemes for evolutive PDEs, Numerical Algorithms, 34 (2003), 245-257.

[8] J.-P. Chehab et B. Costa, Time explicit schemes and spatial finite differences splittings, Journal of Scientific Computing, 20, 2 (2004), pp 159-189.

[9] J.-P. Chehab, B. Costa, Multiparameter extensions of iterative processes, Rapports techniques du laboratoire de mathématiques d'Orsay, RT-02-02, 2002.

[10] C. Wang, J.-G. Liu, Analysis of finite difference schemes for unsteady Navier-Stokes equations in vorticity formulation, Numer. Math. (2002) 91: 543-576

[11] B. Costa, *Time marching techniques for the nonlinear Galerkin method*, Preprint series of the Institute of Applied Mathematics and Scientific Computing, PhD thesis, Bloomington, Indiana, 1998.

[12] B. Costa. L. Dettori, D. Gottlieb and R. Temam, Time marching techniques for the nonlinear Galerkin method, SIAM J. SC. comp., 23, (2001), 1, 46-65.

[13] C.M. Elliott, The Chan-Hilliard Model for the Kinetics of Phase Separation, *in* Mathematical Models for Phase Change Problems, International Series od Numerical Mathematics, Vol. 88, (1989) Birkhäuser.

[14] D. J. Eyre, Unconditionallly Stable One-step Scheme for Gradient Systems, June 1998, unpublished, http://www.math.utah.edu/eyre/research/methods/stable.ps.

[15] Ghia, U.; Ghia, K. N.; Shin, C. T., High-Re solutions for incompressible flow using the Navier-Stokes equations and a multigrid method, Journal of Computational Physics (ISSN 0021-9991), vol. 48, Dec. 1982, p. 387-411.

[16] O. Goyon, High-Reynolds number solutions of Navier-Stokes equations using incremental unknowns, Comput. Methods Appl. Mech. Engrg. 130 (1996) 319-335

[17] S. Lele, Compact Difference Schemes with Spectral Like resolution, J. Comp. Phys., 103, (1992), 16-42

[18] Ming Li, Tao Tang, Bengt Fornberg, A compact fourth-order finite difference scheme for the steady incompressible Navier-Stokes equations, Int. J. Num. Meth. Fluids, Volume 20, Issue 10 (1995) 1137-1151

[19] R. Peyret and R. Taylor, *Computational Methods for Fluid Flow*, Springer Series in Computational Physics (Springer, New-York, 1983).

[20] M. Ribot, *Étude théorique de schémas numériques pour les systèmes de réaction-diffusion; application à des équations paraboliques non linéaires et non locales* , Thèse de Doctorat, Université Claude Bernard - Lyon 1, Décembre 2003 (in French).

[21] M. Ribot, M. Schatzman. Stability, convergence and order of the extrapolations of the Residual Smoothing Scheme in energy norm, Confluentes Math. 3 (2011), no. 3, 495521.

[22] J. Shen, X. Yang, Numerical Approximations of Allen-Cahn and Cahn-Hilliard Equations. DCDS, Series A, (28), (2010), pp 1669–1691.

[23] R. Temam, *Infnite-dimensional dynamical systems in mechanics and physics*, 2nd Ed., Springer-Verlag, New York, 1997.

[24] R. Temam, *Navier-Stokes Equations.* North-Holland, Amsterdam (1984). Revised version

E-mail address: `matthieu.brachet@math.univ-metz.fr`
E-mail address: `Jean-Paul.Chehab@u-picardie.fr`