Sección de Matemáticas
Universidad de La Laguna

Romen Santana Benítez

Métodos SDIRK para ecuaciones de Navier-Stokes incompresibles

SDIRK methods for incompressible Navier-Stokes equations

Trabajo Fin de Grado
Grado en Matemáticas
La Laguna, Junio de 2021

Dirigido por
Domingo Hernández Abreu

Contenido

Resumen/Abstract v

1. Ecuaciones de Navier-Stokes 1
1.1. Ecuaciones de Euler 1
1.2. Ecuaciones de Navier-Stokes 12

2. Métodos de un paso. Métodos de tipo Runge-Kutta 21
2.1. Métodos de un paso: consistencia, estabilidad y convergencia 21
2.2. Métodos de tipo Runge-Kutta 26
2.3. Resolución de la ecuación de etapas en métodos Runge-Kutta implícitos. Métodos DIRK. 29
2.4. Estabilidad absoluta lineal de los métodos Runge-Kutta 32

3. Métodos de proyección para ecuaciones de Navier-Stokes incompresibles 43
3.1. Método de Chorin con Euler Implícito 44
3.2. Método de Chorin con métodos SDIRK 46
3.3. Ilustración numérica 47

Bibliografía 53

Poster 55

Resumen · Abstract

Resumen

Las ecuaciones de Navier-Stokes son unas ecuaciones en derivadas parciales muy importantes tanto en la física como en las matemáticas, pues permiten describir el movimiento para cualquier fluido en el plano y el espacio. Sin embargo, hoy en día se sigue sin poder demostrar o refutar que exista una única solución definida en el espacio para todo tiempo, por lo que sigue siendo un problema abierto de gran interés.

En este trabajo vamos a deducir dichas ecuaciones partiendo de tres principios fundamentales de la mecánica de fluidos: la ley de la conservación de la masa, la segunda ley de Newton (conservación del momento lineal) y la ley de la conservación de la energía. Seguidamente procederemos a introducir los métodos numéricos de tipo Runge-Kutta para ecuaciones diferenciales ordinarias, estudiando su consistencia, estabilidad y convergencia. Esto será necesario pues finalizaremos implementando lo que se conoce como métodos de proyección, concretamente el método de Chorin, considerando métodos de tipo Runge-Kutta simplemente diagonalmente implícitos (SDIRK). Dichos métodos numéricos nos permitirán aproximar la solución de las ecuaciones de Navier-Stokes para fluidos incompresibles. Por último, se ilustrarán varios ejemplos de la implementación del método de Chorin, la cual se ha realizado con la ayuda de Python y el software FEniCS.

Palabras clave: *Ecuaciones de Navier-Stokes – fluidos incompresibles – métodos Runge-Kutta – SDIRK – método de proyección de Chorin.*

Abstract

The Navier-Stokes partial differential equations are very important both in physics and mathematics, as they allow us to describe motion for any fluid in the plane or space. However, today it is still not possible to prove or refute that exists a global solution defined in space for all time, so it remains an open problem of great interest. In this work we are going to deduce these equations starting from three fundamental principles of fluid mechanics: conservation of mass, Newton's second law (balance of momentum) and conservation of energy. Then we will proceed to introduce the Runge-Kutta numerical methods for ordinary differential equations, studying their consistency, stability and convergence. This will be necessary because we will end up implementing what is known as projection methods, specifically the Chorin method, considering simply diagonally implicit Runge-Kutta methods (SDIRK). These numerical methods will allow us to approximate the solution of the Navier-Stokes equations for incompressible fluids. Finally, several examples of the Chorin's method implementation will be illustrated, which has been carried out with the help of Python and the FEniCS software.

Keywords: *Navier-Stokes equations – incompressible fluids – Runge-Kutta methods – SDIRK – Chorin's proyection method*

1

Ecuaciones de Navier-Stokes

En este primer capítulo se deducirán las ecuaciones de Navier-Stokes a partir del desarrollo de las ecuaciones básicas de la mecánica de fluidos. Estas ecuaciones provienen de las leyes de conservación de la masa, energía y momento. Se comenzará con las supuestos más simples hasta llegar a lo que se conoce como las ecuaciones de Euler para un fluido incompresible. Esto se generalizará más adelante para contemplar el efecto de viscosidad que surge del transporte molecular del movimiento. Para el desarrollo de este capítulo nos hemos basado principalmente en [5].

1.1. Ecuaciones de Euler

Sea D una región en un espacio bi- o tridimensional rellena por un fluido, nuestro objetivo es describir el movimiento de dicho fluido. Sea $\mathbf{x} \in D$ un punto en D y consideremos la partícula del fluido moviéndose por $\mathbf{x}$ en un tiempo t. Usando las coordenadas euclídeas en el espacio, escribimos $\mathbf{x}=(x, y, z)$. Imaginemos una partícula (piense en una partícula de polvo suspendida en el aire) en el fluido y sea $\mathbf{u}(\mathbf{x},\mathrm{t})$ la velocidad de la partícula del fluido que se está moviendo por $\mathbf{x}$ en un tiempo t. Por lo tanto, para cada tiempo fijo, $\mathbf{u}$ es un campo de direcciones en D y representa el campo de velocidades del fluido.

Para cada tiempo t, supongamos que el fluido tiene una densidad de masa bien definida $\rho(\mathbf{x}, t)$. Luego, si W es una subregión de D, la masa del fluido en W en un tiempo t viene dada por

$$m(W, t) = \int_W \rho(\mathbf{x}, t)\ dV,$$

donde dV es el elemento diferencial de área en el plano o de volumen en el espacio.

En lo que sigue asumiremos que las funciones $\mathbf{u}$ y ρ (y otras que serán introducidas más adelante) son lo suficientemente regulares como para que los teoremas

del cálculo puedan ser aplicados con validez.

La deducción de las ecuaciones se basa en tres principios básicos:

I Ley de conservación de la masa: la masa no se crea ni se destruye.
II Segunda ley de Newton: la variación de momento de una parte del fluido es igual a la fuerza que se le aplica.
III Ley de conservación de la energía: la energía no se crea ni se destruye.

Tratemos estos tres principios sucesivamente.

I Conservación de la masa

Sea W una subregión fija de D. La variación de la masa en W es

$$\frac{d}{dt}m(W,t) = \frac{d}{dt}\int_W \rho(\mathbf{x},t)\ dV = \int_W \frac{\partial\rho}{\partial t}(\mathbf{x},t)\ dV.$$

Denotemos por ∂W la frontera de W, la cual se asume suficientemente regular; denotemos por $\mathbf{n}$ el vector normal unitario exterior definido en los puntos de ∂W; y denotemos por dA el elemento diferencial del área en ∂W. La tasa de flujo volumétrico por unidad de área a través de ∂W es $\mathbf{u}\cdot\mathbf{n}$ y la tasa de flujo másico por unidad de área es $\rho\mathbf{u}\cdot\mathbf{n}$.

El principio de conservación de la masa se puede enunciar de forma más precisa como sigue: La variación de masa en W es igual a la tasa a la que la masa cruza ∂W con dirección hacia el interior de W, es decir,

$$\frac{d}{dt}\int_W \rho\ dV = -\int_{\partial W} \rho\mathbf{u}\cdot n\ dA. \tag{1.1}$$

Esta es la forma integral para la ley de la conservación de la masa. Por el teorema de la divergencia, esto es equivalente a

$$\int_W \left[\frac{\partial\rho}{\partial t} + \operatorname{div}(\rho\mathbf{u})\right] dV = 0.$$

Como esto se cumple para todo W, por regularidad es equivalente a

$$\frac{\partial\rho}{\partial t} + \operatorname{div}(\rho\mathbf{u}) = 0. \tag{1.2}$$

La ecuación (1.2) es la forma diferencial de la ley de la conservación de la masa, también conocida como ecuación de continuidad.

Nota 1.1 Si ρ y $\mathbf{u}$ no son lo suficientemente regulares para justificar los pasos que llevan a la forma diferencial de la conservación de la masa, se tendría que

usar la forma integral.

II Conservación del momento lineal

Sea $\mathbf{x}(t) = (x(t), y(t), z(t))^T$ la trayectoria que sigue una partícula del fluido, entonces el campo de direcciones de la velocidad viene dado por

$$\mathbf{u}(x(t), y(t), z(t), t) = (\dot{x}(t), \dot{y}(t), \dot{z}(t))^T,$$

esto es,

$$\mathbf{u}(\mathbf{x}(t), t) = \frac{d\mathbf{x}}{dt}(t).$$

Recordar que usamos coordenadas euclídeas en el espacio (suprimiremos z para fluidos en el plano).

La aceleración de una partícula del fluido viene dada por

$$\mathbf{a}(t) = \frac{d^2}{dt^2}\mathbf{x}(t) = \frac{d}{dt}\mathbf{u}(x(t), y(t), z(t), t).$$

Aplicando la regla de la cadena, tenemos que

$$\mathbf{a}(t) = \frac{\partial \mathbf{u}}{\partial x}\dot{x} + \frac{\partial \mathbf{u}}{\partial y}\dot{y} + \frac{\partial \mathbf{u}}{\partial z}\dot{z} + \frac{\partial \mathbf{u}}{\partial t}.$$

Usando la notación

$$\mathbf{u}_x = \frac{\partial \mathbf{u}}{\partial x},\ \mathbf{u}_t = \frac{\partial \mathbf{u}}{\partial t},\ \text{etc.},\quad \mathbf{u} = (u, v, w)^T$$

obtenemos

$$\mathbf{a}(t) = u\mathbf{u}_x + v\mathbf{u}_y + w\mathbf{u}_z + \mathbf{u}_t,$$

que también se puede escribir como

$$\mathbf{a}(t) = \partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla)\mathbf{u},$$

donde

$$\partial_t \mathbf{u} = \frac{\partial \mathbf{u}}{\partial t} \quad \text{y} \quad \mathbf{u} \cdot \nabla = u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}.$$

Llamamos la derivada material al operador:

$$\frac{D}{Dt} = \partial_t + \mathbf{u} \cdot \nabla,$$

que tiene en cuenta el hecho de que el fluido se está moviendo y que la posición de las partículas del fluido varía con el tiempo. En efecto, si $f(x, y, z, t)$ es una

función de posición y tiempo (escalar o vectorial), por la regla de la cadena se tiene que

$$\frac{d}{dt}f(x(t),y(t),z(t),t) = \partial_t f + (\mathbf{u}\cdot\nabla)f = \frac{Df}{Dt}(x(t),y(t),z(t),t). \tag{1.3}$$

Observemos que si $f = (f_1, f_2, f_3)^T$ es vectorial entonces

$$(\mathbf{u}\cdot\nabla)f = ((\mathbf{u}\cdot\nabla)f_1, (\mathbf{u}\cdot\nabla)f_2, (\mathbf{u}\cdot\nabla)f_3)^T = J_f\cdot\mathbf{u},$$

donde J_f es la matriz Jacobiana de f.

Para cualquier continuo, las fuerzas que actúan sobre el material son de dos tipos. Primero, están las fuerzas de tensión, donde el material es afectado por las fuerzas a través de su superficie. Segundo, están las fuerzas externas, como la gravedad o campos magnéticos, que ejercen una fuerza por unidad de volumen al continuo.

Definición 1.1 Se dice que un fluido es ideal si cumple que para cualquier movimiento del fluido existe una función $p(\mathbf{x},t)$ llamada presión tal que si S es una superficie en el fluido con un vector normal exterior fijo y unitario $\mathbf{n}$, la fuerza de tensión ejercida a través de la superficie S por unidad de área en $\mathbf{x}\in S$ en un tiempo t es $p(\mathbf{x},t)\mathbf{n}$. Es decir, para fluidos ideales la fuerza a través de S por unidad de áres es $p(\mathbf{x},t)\mathbf{n}$.

Nota 1.2 Nótese que la fuerza es en la dirección $\mathbf{n}$ y que la fuerza actúa ortogonalmente a la superficie S; por lo tanto, no hay fuerzas tangenciales. Intuitivamente, la ausencia de fuerzas tangenciales implica que no hay posibilidad de que empiece una rotación en el fluido, o, si existiese en un principio, que parase. Por ello los fluidos ideales tienen más interés teórico que práctico ya que hay abundancia de rotación en los fluidos reales (cerca de los remos de un bote, tornados, etc.).

Sea W una región en el fluido para un determinado tiempo t. La fuerza total ejercida sobre el fluido en W por medio de la tensión en su frontera es

$$\mathbf{S}_{\partial W} = -\int_{\partial W} p\mathbf{n}\, dA$$

(es negativa pues $\mathbf{n}$ apunta hacia el exterior). El siguiente resultado representa una versión escalar del teorema de la divergencia.

Lema 1.1

$$\mathbf{S}_{\partial W} = -\int_W \nabla p\, dV$$

Demostración. Sea $\mathbf{e}$ un vector fijo en el espacio. Por el teorema de la divergencia tenemos

$$\mathbf{e} \cdot \mathbf{S}_{\partial W} = -\int_{\partial W} p\mathbf{e} \cdot \mathbf{n}\, dA = -\int_W \operatorname{div}(p\mathbf{e})\, dV = -\int_W (\nabla p) \cdot \mathbf{e}\, dV.$$

Puesto que $\mathbf{e}$ es arbitrario se concluye la prueba.

■

Si $\mathbf{b}(\mathbf{x},t)$ denota la fuerza externa por unidad de masa dada, entonces la fuerza total externa sobre una subregión W es

$$\mathbf{B} = \int_W \rho\mathbf{b}\, dV.$$

Luego, en cualquier región del fluido, la fuerza por unidad de volumen es $-\nabla p + \rho\mathbf{b}$. Por la segunda ley de Newton (fuerza = masa × aceleración) deduciremos la forma diferencial de la ley del balance de momento lineal:

$$\rho\frac{D\mathbf{u}}{Dt} = -\nabla p + \rho\mathbf{b}. \tag{1.4}$$

Ahora procederemos a estudiar la forma integral. Para hacerlo necesitamos introducir algunos conceptos que nos serán útiles. Como hemos hecho previamente, denotemos por D a la región donde el fluido se está moviendo. Sea $\mathbf{x}\in D$, escribiremos $\varphi(\mathbf{x},t)$ para la trayectoria que sigue la partícula que está en la posición $\mathbf{x}$ en tiempo $t=0$. Asumiremos que φ es inversible y lo suficientemente regular para que las siguientes manipulaciones sean legítimas. Denotemos por φ_t a la aplicación $\mathbf{x} \mapsto \varphi(\mathbf{x},t)$; esto es, para un tiempo t, esta aplicación asigna a cada partícula del fluido su posición inicial en t=0 a su posición actual en tiempo t (de modo que φ_0 es la aplicación identidad). Llamaremos a φ el flujo del fluido. Si W es una región en D, entonces $\varphi_t(W) = W_t$ representa el volumen W moviéndose con el fluido.

La forma integral "original"del balance de momento establece que

$$\frac{d}{dt}\int_{W_t} \rho\mathbf{u}\, dV = S_{\partial W_t} + \int_{W_t} \rho\mathbf{b}\, dV, \tag{1.5}$$

esto es, la variación de momento de cualquier parte del fluido en movimiento es igual al total de las fuerzas (suma de las fuerzas de tensión y externas) que actúan sobre él.

Teorema 1.1. *La forma diferencial del balance de momento (1.4) y la forma integral (1.5) son equivalentes.*

Demostración. Para demostrar esto partimos de la forma integral (1.5). Aplicando el teorema del cambio de variable para $\mathbf{x}$ tenemos que

$$\frac{d}{dt}\int_{W_t}\rho\mathbf{u}\,dV = \frac{d}{dt}\int_{W}(\rho\mathbf{u})(\varphi(\mathbf{x},t))J(\mathbf{x},t)\,dV,$$

donde $J(\mathbf{x},t)$ es el determinante Jacobiano del flujo φ_t. Aquí asumimos que $J(\mathbf{x},t)\geq 0$, es decir que el flujo φ_t conserva la orientación en el fluido. Ahora, aplicando la regla de Leibniz

$$\frac{d}{dt}\int_{W}(\rho\mathbf{u})(\varphi(\mathbf{x},t))J(\mathbf{x},t)\,dV = \int_{W}\frac{\partial}{\partial t}\left[(\rho\mathbf{u})(\varphi(\mathbf{x},t))J(\mathbf{x},t))\right]\,dV,$$

y por la regla de la cadena

$$\begin{aligned}\int_{W}\frac{\partial}{\partial t}\left[(\rho\mathbf{u})(\varphi(\mathbf{x},t))J(\mathbf{x},t)\right]\,dV &= \int_{W}\frac{\partial}{\partial t}\left[(\rho\mathbf{u})(\varphi(\mathbf{x},t))\right]J(\mathbf{x},t)\\ &\quad+\frac{\partial}{\partial t}\left[J(\mathbf{x},t)\right](\rho\mathbf{u})(\varphi(\mathbf{x},t))\,dV.\end{aligned}$$

Sabemos por (1.3) que

$$\frac{\partial}{\partial t}(\rho\mathbf{u})(\varphi(\mathbf{x},t),t) = \left(\frac{D}{Dt}(\rho\mathbf{u})\right)(\varphi(\mathbf{x},t),t),$$

y para derivar $J(\mathbf{x},t)$ aplicamos el *Lema 1.2* (cuya demostración posponemos), en virtud del cual

$$\frac{\partial}{\partial t}J(\mathbf{x},t) = J(\mathbf{x},t)[\text{div }\mathbf{u}(\varphi(\mathbf{x},t),t)].$$

Así pues, llegamos a

$$\frac{d}{dt}\int_{W_t}\rho\mathbf{u}\,dV = \int_{W}\left\{\left(\frac{D}{Dt}(\rho\mathbf{u})\right)(\varphi(\mathbf{x},t),t)+(\rho\mathbf{u})(\text{div }\mathbf{u})(\varphi(\mathbf{x},t),t)\right\}\cdot J(\mathbf{x},t)\,dV.$$

Deshaciendo el cambio de variable y aplicando de nuevo la regla de la cadena se tiene que

$$\begin{aligned}\frac{d}{dt}\int_{W_t}\rho\mathbf{u}\,dV &= \int_{W_t}\left\{\frac{D}{Dt}(\rho\mathbf{u})+(\rho\text{ div }\mathbf{u})\mathbf{u}\right\}\,dV\\ &= \int_{W_t}\left\{\frac{D\rho}{Dt}\mathbf{u}+\rho\frac{D\mathbf{u}}{Dt}+(\rho\text{ div }\mathbf{u})\mathbf{u}\right\}\,dV.\end{aligned}$$

Observando que

$$\frac{D\rho}{Dt}+\rho\cdot\text{div }\mathbf{u} = \rho_t+(\mathbf{u}\cdot\nabla)\rho+\rho\cdot\text{div }\mathbf{u} = \rho_t+\text{div}(\rho\cdot\mathbf{u}),$$

por la ley de la conservación de la masa (1.2) llegamos a

$$\frac{d}{dt}\int_{W_t} \rho\mathbf{u}\ dV = \int_{W_t} \rho\frac{D\mathbf{u}}{Dt}\ dV.$$

Finalmente, podemos afirmar partiendo de (1.5) que

$$\frac{d}{dt}\int_{W_t} \rho\mathbf{u}\ dV = \int_{W_t} \rho\frac{D\mathbf{u}}{Dt}\ dV = -\int_{W_t} \nabla p\ dV + \int_{W_t} \rho\mathbf{b}\ dV.$$

Entonces si W es arbitrario, y por lo tanto W_t lo es, y los integrandos son continuos, obtendríamos que

$$\rho\frac{D\mathbf{u}}{Dt} = -\nabla p + \rho\mathbf{b},$$

esto es, (1.4). Por otra parte, es claro que (1.4) implica (1.5).

■

El argumento de la prueba del *Teorema 1.1* nos conduce al siguiente resultado.

Corolario 1.1 (Teorema del Transporte) Para cualquier función f que dependa de $\mathbf{x}$ y t, se tiene que

$$\frac{d}{dt}\int_{W_t} \rho f\ dV = \int_{W_t} \rho\frac{Df}{Dt}\ dV.$$

En particular,

$$\frac{d}{dt}\int_{W_t} f\ dV = \int_{W_t} \frac{Df}{Dt}\ dV.$$

Lema 1.2 Sea $J(\mathbf{x},t)$ el determinante Jacobiano del flujo φ_t. Entonces

$$\frac{\partial}{\partial t}J(\mathbf{x},t) = J(\mathbf{x},t)[\text{div } \mathbf{u}(\varphi(\mathbf{x},t),t)].$$

Demostración. Escribimos las componentes de φ como $\xi(\mathbf{x},t), \eta(\mathbf{x},t), \zeta(\mathbf{x},t)$. Primero, observar que

$$\frac{\partial}{\partial t}\varphi(\mathbf{x},t) = \mathbf{u}(\varphi(\mathbf{x},t),t),$$

por la definición del campo de velocidades del fluido.
El determinante J se puede derivar recordando que el determinante de una matriz es multilineal en las columnas (o en las filas). Luego para $\mathbf{x}$ fijo, tenemos

$$\frac{\partial}{\partial t}J = \begin{vmatrix} \frac{\partial}{\partial t}\frac{\partial \xi}{\partial x} & \frac{\partial \eta}{\partial x} & \frac{\partial \zeta}{\partial x} \\ \frac{\partial}{\partial t}\frac{\partial \xi}{\partial y} & \frac{\partial \eta}{\partial y} & \frac{\partial \zeta}{\partial y} \\ \frac{\partial}{\partial t}\frac{\partial \xi}{\partial z} & \frac{\partial \eta}{\partial z} & \frac{\partial \zeta}{\partial z} \end{vmatrix} + \begin{vmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial}{\partial t}\frac{\partial \eta}{\partial x} & \frac{\partial \zeta}{\partial x} \\ \frac{\partial \xi}{\partial y} & \frac{\partial}{\partial t}\frac{\partial \eta}{\partial y} & \frac{\partial \zeta}{\partial y} \\ \frac{\partial \xi}{\partial z} & \frac{\partial}{\partial t}\frac{\partial \eta}{\partial z} & \frac{\partial \zeta}{\partial z} \end{vmatrix} + \begin{vmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \eta}{\partial x} & \frac{\partial}{\partial t}\frac{\partial \zeta}{\partial x} \\ \frac{\partial \xi}{\partial y} & \frac{\partial \eta}{\partial y} & \frac{\partial}{\partial t}\frac{\partial \zeta}{\partial y} \\ \frac{\partial \xi}{\partial z} & \frac{\partial \eta}{\partial z} & \frac{\partial}{\partial t}\frac{\partial \zeta}{\partial z} \end{vmatrix}.$$

Ahora escribimos

$$\begin{aligned} \frac{\partial}{\partial t}\frac{\partial \xi}{\partial x} &= \frac{\partial}{\partial x}\frac{\partial \xi}{\partial t} = \frac{\partial}{\partial x}u(\varphi(\mathbf{x},t),t), \\ \frac{\partial}{\partial t}\frac{\partial \xi}{\partial y} &= \frac{\partial}{\partial y}\frac{\partial \xi}{\partial t} = \frac{\partial}{\partial y}u(\varphi(\mathbf{x},t),t), \\ &\vdots \\ \frac{\partial}{\partial t}\frac{\partial \zeta}{\partial z} &= \frac{\partial}{\partial z}\frac{\partial \zeta}{\partial t} = \frac{\partial}{\partial z}w(\varphi(\mathbf{x},t),t). \end{aligned}$$

Las componentes u, v y w en esta expresión son funciones de x, y, z evaluadas en $\varphi(\mathbf{x}, t)$; por lo tanto

$$\begin{aligned} \frac{\partial}{\partial x}u(\varphi(\mathbf{x},t),t) &= \frac{\partial u}{\partial \xi}\frac{\partial \xi}{\partial x} + \frac{\partial u}{\partial \eta}\frac{\partial \eta}{\partial x} + \frac{\partial u}{\partial \zeta}\frac{\partial \zeta}{\partial x}, \\ \frac{\partial}{\partial y}u(\varphi(\mathbf{x},t),t) &= \frac{\partial u}{\partial \xi}\frac{\partial \xi}{\partial y} + \frac{\partial u}{\partial \eta}\frac{\partial \eta}{\partial y} + \frac{\partial u}{\partial \zeta}\frac{\partial \zeta}{\partial y}, \\ &\vdots \\ \frac{\partial}{\partial z}w(\varphi(\mathbf{x},t),t) &= \frac{\partial w}{\partial \xi}\frac{\partial \xi}{\partial z} + \frac{\partial w}{\partial \eta}\frac{\partial \eta}{\partial z} + \frac{\partial w}{\partial \zeta}\frac{\partial \zeta}{\partial z}. \end{aligned}$$

Sustituyendo esto en la expresión $\dfrac{\partial}{\partial t}J$, obtenemos que

$$\frac{\partial}{\partial t}J = \frac{\partial u}{\partial \xi}J + \frac{\partial v}{\partial \eta}J + \frac{\partial w}{\partial \zeta}J = (\text{div } \mathbf{u})J.$$

■

Definición 1.1 En términos de la notación introducida anteriormente, un fluido se dice incompresible si para cualquier subregión W, se tiene que

$$\text{volumen}(W_t) = \int_{W_t} dV \text{ es constante en t.}$$

El *Lema 1.2* es también útil para entender la incompresibilidad en fluidos, según refleja el siguiente resultado:

Teorema 1.2. *Las siguientes condiciones son equivalentes:*

i) *un fluido es incompresible*
ii) div $\mathbf{u} = 0$
iii) $J \equiv 1$

Demostración. i)⇒ii)
Usando (i) y el *Lema 1.2*:

$$0 = \frac{d}{dt}\int_{W_t} dV = \frac{d}{dt}\int_W J\ dV = \int_W (\text{div } \mathbf{u}) J\ dV = \int_{W_t} (\text{div } \mathbf{u})\ dV.$$

Puesto que esto se verifica para toda región W, se deduce que div $\mathbf{u} = 0$.
ii)⇒iii)
Tenemos por el *Lema 1.2* y ii) que

$$\begin{cases} \dfrac{\partial J}{\partial t} = (\text{div } \mathbf{u})J = 0 \\ J(\mathbf{x}, 0) = \det I = 1, \text{ pues } \varphi(\mathbf{x}, 0) = \mathbf{x} \end{cases}$$

Esto implica que $J \equiv 1$.
iii)⇒i)
Nuevamente por el *Lema 1.2* y usando iii):

$$\int_{W_t} dV = \int_W J\ dV = \int_W dV$$

lo que implica que $\int_{W_t} dV$ es constante en t.

■

Nota 1.3 De la ecuación de continuidad (1.2)

$$\frac{\partial \rho}{\partial t} + \text{div}(\rho \mathbf{u}) = 0, \text{ es decir, } \frac{D\rho}{Dt} + \rho \text{ div } \mathbf{u} = 0,$$

y del hecho de que $\rho > 0$, vemos que un fluido es incompresible si y solo si $D\rho/Dt = 0$, lo que equivale a decir que la densidad de masa es constante siguiendo la trayectoria del fluido. Si el fluido es homogéneo, es decir, ρ es constante en $\mathbf{x}$, también se obtiene que el fluido es incompresible si y solo si ρ es constante en el tiempo.

III Conservación de la energía

Hasta ahora hemos desarrollado las ecuaciones

$$\frac{D\rho}{Dt} + \rho \text{ div } \mathbf{u} = 0 \quad \text{(conservación de la masa)}$$

y

$$\rho\frac{D\mathbf{u}}{Dt} = -\nabla p + \rho\mathbf{b} \text{ (balance del momento lineal).}$$

Esto son cuatro ecuaciones si trabajamos en un espacio tridimensional (o $n+1$ ecuaciones si trabajamos en un espacio de dimensión n), porque la ecuación de balance de momento es una ecuación vectorial compuesta por tres ecuaciones escalares. Sin embargo, tenemos cinco incógnitas: $\mathbf{u}$, ρ y p. Por lo tanto, para especificar el movimiento de un fluido se necesita una ecuación más. La conservación de la energía va a suministrar una ecuación adicional.

Para un fluido moviéndose en un dominio D, con un campo de velocidades $\mathbf{u}$, la energía cinética contenida en una región $W \subset D$ es:

$$E_{\text{cinética}} = \frac{1}{2}\int_W \rho \left\|\mathbf{u}\right\|^2 \, dV$$

donde $\left\|\mathbf{u}\right\|^2 = (u^2+v^2+w^2)$ es el cuadrado del módulo de la velocidad $\mathbf{u}$. La variación de la energía cinética de una porción en movimiento W_t del fluido se calcula usando el teorema del transporte como sigue:

$$\begin{aligned}
\frac{d}{dt}E_{\text{cinética}} &= \frac{d}{dt}\left[\frac{1}{2}\int_{W_t} \rho \left\|u\right\|^2 \, dV\right] \\
&= \frac{1}{2}\int_{W_t} \rho\frac{D\left\|u\right\|^2}{Dt} \, dV \\
&= \int_{W_t} \rho\left(\mathbf{u}\cdot\left(\frac{\partial\mathbf{u}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{u}\right)\right) dV \\
&= \int_{W_t} \rho\left(\mathbf{u}\cdot\left(\frac{D\mathbf{u}}{Dt}\right)\right) dV.
\end{aligned}$$

donde la tercera y cuarta igualdad son consecuencia de que

$$\begin{aligned}
\frac{1}{2}\frac{D}{Dt}\left\|\mathbf{u}\right\|^2 &= \frac{1}{2}\frac{\partial}{\partial t}(u^2+v^2+w^2) + \frac{1}{2}\left(u\frac{\partial}{\partial x}(u^2+v^2+w^2)\right. \\
&\quad \left. + v\frac{\partial}{\partial y}(u^2+v^2+w^2) + w\frac{\partial}{\partial z}(u^2+v^2+w^2)\right) \\
&= u\frac{\partial u}{\partial t} + v\frac{\partial v}{\partial t} + w\frac{\partial w}{\partial t} + u\left(u\frac{\partial u}{\partial x} + v\frac{\partial v}{\partial x} + w\frac{\partial w}{\partial x}\right) \\
&\quad + v\left(u\frac{\partial u}{\partial y} + v\frac{\partial v}{\partial y} + w\frac{\partial w}{\partial y}\right) + w\left(u\frac{\partial u}{\partial z} + v\frac{\partial v}{\partial z} + w\frac{\partial w}{\partial z}\right) \\
&= \mathbf{u}\cdot\frac{\partial\mathbf{u}}{\partial t} + \mathbf{u}\cdot((\mathbf{u}\cdot\nabla)\mathbf{u}) = \mathbf{u}\cdot\frac{D\mathbf{u}}{Dt}.
\end{aligned}$$

Asumiremos que la energía total del fluido se puede escribir como

$$E_{\text{total}} = E_{\text{cinética}} + E_{\text{interna}}$$

donde E_{interna} es la energía interna que denotaremos por ϵ y proviene de fuentes como el potencial intermolecular o las vibraciones moleculares internas.
Si se añade energía al fluido o le permitimos realizar trabajo, la energía total cambiará. La variación de la energía total de una porción del fluido W_t es igual al trabajo realizado sobre él, esto es, el trabajo realizado por fuerzas externas y la presión

$$\begin{aligned}\frac{d}{dt}E_{\text{total}} &= \frac{d}{dt}\int_{W_t}(\frac{1}{2}\rho\left\|\mathbf{u}\right\|^2 + \rho\epsilon)\ dV \\ &= \int_{W_t}\rho\mathbf{u}\cdot\mathbf{b}\ dV - \int_{\partial W_t} p\mathbf{u}\cdot\mathbf{n}\ dA.\end{aligned}$$

Una discusión general de la conservación de la energía requiere un mayor conocimiento de termodinámica del que necesitamos. Nosotros nos limitaremos al ejemplo de conservación de energía para fluidos incompresibles que es lo que nos concierne en este trabajo.

Ecuaciones de Euler para fluidos incompresibles ideales

Sabemos que, en general, la variación de energía cinética se expresa como

$$\begin{aligned}\frac{d}{dt}E_{\text{cinética}} &= \int_{W_t}\rho\left(\mathbf{u}\cdot\frac{D\mathbf{u}}{Dt}\right)dV \\ &= \int_{W_t}(-\nabla p\cdot\mathbf{u} + \rho\mathbf{b}\cdot\mathbf{u})dV,\end{aligned}$$

a consecuencia de la ecuación del balance del momento (1.4).
Por otro lado, hemos visto que

$$\frac{d}{dt}E_{\text{total}} = \int_{W_t}\rho\mathbf{u}\cdot\mathbf{b}\ dV - \int_{\partial W_t} p\mathbf{u}\cdot\mathbf{n}\ dA.$$

Luego si asumimos $\text{div}(\mathbf{u}) = 0$ y aplicamos el teorema de la divergencia obtenemos que:

$$\frac{d}{dt}E_{\text{cinética}} = \int_{W_t}(-\text{div}(p\mathbf{u}) + \rho\mathbf{u}\cdot\mathbf{b})dV = \frac{d}{dt}E_{\text{total}}$$

Recíprocamente si asumimos que la energía cinética coincide con la energía total y que $p \neq 0$ entonces:

$$\frac{d}{dt}E_{\text{cinética}} = \frac{d}{dt}E_{\text{total}} \Rightarrow \int_{W_t} p\ \text{div}\ \mathbf{u}\ dV = 0,\ \forall W, \forall t \Rightarrow \text{div}\ \mathbf{u}\ = 0.$$

Este argumento lleva a que si $E_{\text{total}} = E_{\text{cinética}}$ entonces el fluido debe ser incompresible (siempre que $p \neq 0$). En resumen, para el caso incompresible, las Ecuaciones de Euler para fluidos incompresibles ideales son:

$$\begin{cases} \rho \dfrac{D\mathbf{u}}{Dt} = -\nabla p + \rho \mathbf{b} \\ \dfrac{D\rho}{Dt} = 0 \\ \text{div } \mathbf{u} = 0 \end{cases}$$

que se dotan usualmente con las condiciones de frontera homogéneas

$$\mathbf{u} \cdot \mathbf{n} = 0 \ \text{ en } \partial D.$$

Nota 1.4 En ausencia de fuerzas externas, esto es, $\mathbf{b} \equiv 0$, tendríamos que

$$\frac{d}{dt} E_{\text{cinética}} = \int_D -\text{div}(p\mathbf{u}) \, dV = \int_{\partial D} p(\mathbf{u} \cdot \mathbf{n}) \, dA,$$

y como $\mathbf{u} \cdot \mathbf{n} = 0 \ \text{ en } \partial D$, resulta que

$$\frac{d}{dt} E_{\text{cinética}} = 0.$$

Luego, para fluidos incompresibles ideales la variación de energía cinética en toda la región D es nula.

1.2. Ecuaciones de Navier-Stokes

En §**1.1** definimos un fluido ideal como aquel en el que las fuerzas que actúan a través de la superficie S son perpendiculares a ésta. Ahora consideraremos fluidos más generales. De modo que en vez de asumir que la fuerza sobre S por unidad de área es $-p(\mathbf{x}, t)\mathbf{n}$ donde $\mathbf{n}$ es la normal exterior en la frontera S, ahora asumiremos que la fuerza sobre S por unidad de área es

$$-p(\mathbf{x}, t)\mathbf{n} + \boldsymbol{\sigma} \cdot \mathbf{n}$$

donde $\boldsymbol{\sigma}$ es una matriz llamada el tensor de tensión, sobre la cual deberemos hacer algunas suposiciones. Una de las características es que $\boldsymbol{\sigma} \cdot \mathbf{n}$ no tiene que ser necesariamente paralelo a $\mathbf{n}$, por lo que esta separación es un tanto ambigua. Como hemos visto anteriormente, la segunda ley de Newton establece que la variación de momento de una porción del fluido en movimiento W_t es igual a la resultante de fuerzas que actúan sobre él (balance del momento lineal):

$$\frac{d}{dt} \int_{W_t} \rho \mathbf{u} \, dV = -\int_{\partial W_t} (p \cdot \mathbf{n} - \boldsymbol{\sigma} \cdot \mathbf{n}) dA$$

(comparar con (1.5) en §**1.1**). En esta sección asumimos por simplicidad en la exposición ausencia de fuerzas externas, esto es, $\mathbf{b} \equiv 0$. Vemos que $\boldsymbol{\sigma}$ modifica el transporte de momento a través de la superficie de W_t.

Nota 1.5 Podríamos preguntarnos por qué las fuerzas que actúan sobre S deberían ser una función que depende linealmente de $\mathbf{n}$. De hecho, si asumimos que la fuerza es una función continua de $\mathbf{n}$, usando el balance de momento, se podría demostrar que es lineal en $\mathbf{n}$. Este resultado se conoce como el *Teorema de Cauchy* [5].

Las condiciones que asumiremos para $\boldsymbol{\sigma}$ son las siguientes:

1. $\boldsymbol{\sigma}$ *depende linealmente del gradiente de velocidad* $\nabla\mathbf{u}$, esto es
$$\boldsymbol{\sigma} = \mathrm{a}(\mathbf{x},t)\cdot\mathbf{I} + \mathrm{b}(\mathbf{x},t)\cdot\nabla\mathbf{u}, \quad \text{con } \nabla\mathbf{u} = \begin{pmatrix} \partial_x u & \partial_y u & \partial_z u \\ \partial_x v & \partial_y v & \partial_z v \\ \partial_x w & \partial_y w & \partial_z w \end{pmatrix}$$
donde $\mathbf{I}$ es la matriz identidad.
2. $\boldsymbol{\sigma}$*es invariante bajo rotaciones rígidas*, esto es, si $\mathbf{U}$ es una matriz ortogonal,
$$\boldsymbol{\sigma}(\mathbf{U}\cdot\nabla\mathbf{u}\cdot\mathbf{U}^{-1}) = \mathbf{U}\cdot\boldsymbol{\sigma}(\nabla\mathbf{u})\cdot\mathbf{U}^{-1}$$
3. $\boldsymbol{\sigma}$ *es simétrica*

Denotemos por $\mathbf{D} = \frac{1}{2}[\nabla\mathbf{u} + (\nabla\mathbf{u})^T]$ y $\mathbf{S} = \frac{1}{2}[\nabla\mathbf{u} - (\nabla\mathbf{u})^T]$ a la parte simétrica y antisimétrica de $\nabla\mathbf{u}$, respectivamente, de modo que $\nabla\mathbf{u} = \mathbf{D} + \mathbf{S}$; como $\boldsymbol{\sigma}$ es simétrica, por la propiedad 1 sigue que $\boldsymbol{\sigma}$ depende solo de $\mathbf{D}$. Esto se debe a:

$$\begin{aligned} \boldsymbol{\sigma} = \boldsymbol{\sigma}^T &\Rightarrow \mathrm{a}\cdot\mathbf{I} + \mathrm{b}\cdot\nabla\mathbf{u} = \mathrm{a}\cdot\mathbf{I} + \mathrm{b}\cdot(\nabla\mathbf{u})^T \\ &\Rightarrow \mathrm{b}\cdot[\nabla\mathbf{u} - (\nabla\mathbf{u})^T] = 0 \Rightarrow \mathrm{b}\cdot\mathbf{S} = 0 \\ &\Rightarrow \boldsymbol{\sigma} = \mathrm{a}\cdot\mathbf{I} + \mathrm{b}\cdot[\mathbf{D}+\mathbf{S}] = \mathrm{a}\cdot\mathbf{I} + \mathrm{b}\cdot\mathbf{D}. \end{aligned}$$

Como $\boldsymbol{\sigma}$ es una función que depende linealmente de $\mathbf{D}$ y $\mathbf{D}$ es simétrica, ambas puedes ser diagonalizadas simultáneamente. Luego, los autovalores de $\boldsymbol{\sigma}$ dependerán linealmente de los de $\mathbf{D}$. Por la propiedad 2, estos deben ser también simétricos porque podremos elegir $\mathbf{U}$ para permutar dos autovalores de $\mathbf{D}$ y esto debe permutar los correspondientes autovalores de $\boldsymbol{\sigma}$. Las únicas funciones lineales simétricas que cumplen esto son de la forma

$$\sigma_i = \lambda(d_1 + d_2 + d_3) + 2\mu d_i, \quad i = 1,2,3,$$

donde σ_i son los autovalores de $\boldsymbol{\sigma}$ y d_i los de $\mathbf{D}$. Recordemos que la traza de una matriz es invariante bajo transformaciones ortogonales. Además tenemos que

$$d_1 + d_2 + d_3 = \text{traza de } \mathbf{D} = \text{traza de } \frac{1}{2}[\nabla \mathbf{u} + (\nabla \mathbf{u})^T] = \text{div } \mathbf{u},$$

y usando la propiedad 2 deducimos que

$$\boldsymbol{\sigma} = \lambda(\text{div } \mathbf{u})\mathbf{I} + 2\mu\mathbf{D}. \tag{1.6}$$

Ahora procederemos a la deducción de las ecuaciones de Navier-Stokes, para lo cual asumiremos en lo que sigue que λ y μ son constantes.

Partimos de la ecuación de balance de momento

$$\frac{\partial}{\partial t}\int_{W_t} \rho\mathbf{u}\, dV = \int_{\partial W_t} (-p\cdot\mathbf{n} + \boldsymbol{\sigma}\cdot\mathbf{n})\, dA.$$

Aplicando el teorema del transporte

$$\frac{\partial}{\partial t}\int_{W_t} \rho\mathbf{u}\, dV = \int_{W_t} \rho\frac{D\mathbf{u}}{Dt}\, dV,$$

y por el teorema de la divergencia

$$\int_{\partial W_t} (-p\cdot\mathbf{n} + \boldsymbol{\sigma}\cdot\mathbf{n})\, dA = \int_{W_t} (-\nabla p)\, dV + \int_{\partial W_t} (\boldsymbol{\sigma}\cdot\mathbf{n})\, dA.$$

Además, aplicando nuevamente el teorema de la divergencia

$$\begin{aligned}\int_{\partial W_t} (\boldsymbol{\sigma}\cdot\mathbf{n})\, dA &= \int_{\partial W_t} \lambda(\text{div } \mathbf{u})\cdot\mathbf{n}\, dA + \int_{\partial W_t} 2\mu\mathbf{D}\cdot\mathbf{n}\, dA \\ &= \lambda\int_{W_t} \nabla(\text{div } \mathbf{u})\, dV + \mu\int_{\partial W_t} [\nabla\mathbf{u} + (\nabla\mathbf{u})^T]\cdot\mathbf{n}\, dA \\ &= \lambda\int_{W_t} \nabla(\text{div } \mathbf{u})\, dV + \mu\int_{W_t} [\Delta\mathbf{u} + \nabla(\text{div } \mathbf{u})]\, dV\end{aligned}$$

donde

$$\Delta\mathbf{u} = (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2})\mathbf{u}$$

es el Laplaciano de $\mathbf{u}$. De aquí que

$$\int_{W_t} \rho\frac{D\mathbf{u}}{Dt} dV = \int_{W_t} [-\nabla p + (\lambda+\mu)\nabla(\text{div } \mathbf{u}) + \mu\Delta\mathbf{u}]\, dV.$$

Finalmente, por verificarse esta igualdad para todo W y t, asumiendo regularidad obtenemos que

$$\rho\frac{D\mathbf{u}}{Dt} = -\nabla p + (\lambda+\mu)\nabla(\text{div } \mathbf{u}) + \mu\Delta\mathbf{u}. \tag{1.7}$$

(1.7) junto con la ecuación de continuidad y una ecuación para la energía, describiría completamente el flujo para un fluido viscoso compresible.

En el caso de un fluido homogéneo incompresible ($\rho = \rho_0 =$ constante), el conjunto completo de ecuaciones se convierte en las ecuaciones de Navier-Stokes para fluidos incompresibles,

$$\begin{aligned} \frac{D\mathbf{u}}{Dt} &= -\nabla p' + \nu \Delta \mathbf{u} \\ \text{div } \mathbf{u} &= 0 \end{aligned} \tag{1.8}$$

donde $\nu = \mu/\rho_0$ es el coeficiente de viscosidad cinética, y $p' = p/\rho_0$.
Estas ecuaciones son suplementadas con condiciones de frontera adecuadas. En las ecuaciones de Euler para un fluido ideal se usa $\mathbf{u}\cdot\mathbf{n} = 0$, es decir, el fluido no cruza la frontera pero si puede moverse tangencialmente a ésta. En las ecuaciones de Navier-Stokes, el término $\nu\Delta\mathbf{u}$ eleva el orden de las derivadas que implican a $\mathbf{u}$, lo cual viene acompañado de un aumento del número de condiciones en la frontera. Por ejemplo, para una pared sólida en reposo se le añade la condición de que la velocidad tangencial es también cero (condición *"no-slip"*), luego las condiciones en la frontera serían $\mathbf{u} = \mathbf{0}$ para paredes sólidas en reposo.

Nota 1.6 La necesidad de introducir condiciones de frontera extra proviene del papel que desempeñan para demostrar que las ecuaciones están bien planteadas, es decir, que exista una solución única y que ésta tenga continuidad respecto al dato inicial. Se sabe que, en tres dimensiones, las ecuaciones de Navier-Stokes incompresibles admiten solución regular para tiempos cortos, con dependencia continua respecto al dato inicial. Sin embargo, es un problema abierto en la mecánica de fluidos probar o refutar que las soluciones existen para todo tiempo. En dos dimensiones se conoce que existe solución para todo tiempo, tanto en fluidos no viscosos como viscosos [5].

Número de Reynolds

A continuación, trataremos algunas propiedades de escalamiento en las ecuaciones de Navier-Stokes con el objetivo de introducir un parámetro (el número de Reynolds) que mida el efecto de viscosidad en el fluido.

Para un problema dado, supongamos que L y U son lo que se conoce como una longitud y velocidad características, respectivamente, representando cierta estimación arbitraria de la longitud y velocidad del problema. La elección de L y U determinará una escala temporal $T = L/U$. En vez de trabajar con las variables $\mathbf{x}$, $\mathbf{u}$ y t lo que haremos será reescalarlas:

$$\mathbf{u}' = \frac{\mathbf{u}}{U}, \quad \mathbf{x}' = \frac{\mathbf{x}}{L} \quad \text{y} \quad t' = \frac{t}{T}.$$

La componente en x de las ecuaciones de Navier-Stokes incompresibles en el caso de un fluido homogéneo es

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} = -\frac{1}{\rho_0}\frac{\partial p}{\partial x} + \nu\left[\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right].$$

El cambio de variables nos lleva a

$$\begin{aligned}
\frac{\partial u}{\partial t} &= \frac{\partial u}{\partial t'}\frac{\partial t'}{\partial t} = \frac{1}{T}\frac{\partial u}{\partial t'} = \frac{U^2}{L}\frac{\partial u'}{\partial t'} \\
\frac{\partial u}{\partial x} &= \frac{\partial u}{\partial x'}\frac{\partial x'}{\partial x} = \frac{1}{L}\frac{\partial u}{\partial x'} = \frac{U}{L}\frac{\partial u'}{\partial x'} \\
\frac{\partial p}{\partial x} &= \frac{\partial p}{\partial x'}\frac{\partial x'}{\partial x} = \frac{1}{L}\frac{\partial p}{\partial x'} \\
\frac{\partial^2 u}{\partial x^2} &= \frac{\partial}{\partial x'}\left(\frac{\partial u}{\partial x}\right)\frac{\partial x'}{\partial x} = \frac{1}{L}\frac{\partial}{\partial x'}\left(\frac{\partial u}{\partial x}\right) = \frac{U}{L^2}\frac{\partial^2 u'}{\partial (x')^2}.
\end{aligned}$$

Identidades análogas se obtienen para las componentes en y y z. Si combinamos todas las componentes y dividimos por U^2/L, obtenemos

$$\frac{\partial \mathbf{u}'}{\partial t'} + (\mathbf{u}' \cdot \nabla')\mathbf{u}' = -\nabla p' + \frac{\nu}{LU}\Delta' \mathbf{u}', \tag{1.9}$$

donde $p' = p/(\rho_0 U^2)$. La incompresibilidad se expresa igualmente como

$$\mathbf{div}\ \mathbf{u}' = 0.$$

Las ecuaciones (1.9) son las ecuaciones adimensionales de Navier-Stokes. Se denomina número de Reynolds R al parámetro adimensional

$$R = \frac{LU}{\nu}$$

Dos fluidos con la misma geometría y el mismo número de Reynolds se denominan similares. De forma más precisa, sean $\mathbf{u}_1$ y $\mathbf{u}_2$ las velocidades respectivas de dos flujos en dos regiones D_1 y D_2 que estén relacionadas por un factor de escala λ de tal manera que $L_1 = \lambda L_2$. Si se cumple que

$$\frac{L_1 U_1}{\nu_1} = \frac{L_2 U_2}{\nu_2}, \text{ esto es, } R_1 = R_2$$

entonces los campos de velocidades adimensionales $\mathbf{u}_1'$ y $\mathbf{u}_2'$ satisfacen exactamente la misma ecuación en la misma región. Por lo tanto, podemos concluir

que $\mathbf{u}_1$ se puede obtener mediante un reescalamiento de $\mathbf{u}_2$; en otras palabras, $\mathbf{u}_1$ y $\mathbf{u}_2$ son similares.

Nota 1.7 Serán de especial interés los casos donde R es grande. Resaltar que no se puede afirmar que si "ν es pequeño" entonces los efectos de viscosidad no tendrían importancia, pues tendríamos que tener en cuenta las dimensiones del problema, pero decir que "$1/R$ es pequeño" sí es significativo.

Nota 1.8 En las ecuaciones de Navier-Stokes para un fluido viscoso incompresible

$$\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla)\mathbf{u} = -\nabla p + \frac{1}{R}\Delta \mathbf{u}, \tag{1.10}$$

el término $\frac{1}{R}\Delta\mathbf{u}$ se denomina término de difusión o disipativo y $(\mathbf{u} \cdot \nabla)\mathbf{u}$ el término de transporte o convectivo. Las ecuaciones establecen que el material es transportado sujeto a fuerzas de presión y, al mismo tiempo, disipado.
Como ocurre para un fluido ideal incompresible, la presión p para un fluido viscoso incompresible viene determinada por la ecuación div $\mathbf{u} = 0$. Ahora discutiremos el papel de la presión para un fluido incompresible con mayor profundidad. Usaremos el siguiente teorema de descomposición.

Teorema 1.3 (Descomposición de Helmholtz-Hodge). *Sea D una región en el espacio(o del plano) con frontera regular ∂D. Sea $\mathbf{w} \in C^1$ un campo vectorial en $D \cup \partial D$, entonces existe un campo vectorial $\mathbf{u} \in C^1$ paralelo a la frontera con divergencia cero, es decir, $\mathbf{u} \cdot \mathbf{n} = 0$ en ∂D y* div $\mathbf{u}$=0, *y una función escalar $p \in C^2$, tal que*

$$\mathbf{w} = \mathbf{u} + \nabla p. \tag{1.11}$$

Esta descomposición es única salvo constante aditiva en p.

Demostración. Observamos que si la descomposición (1.11) es posible, entonces div $\mathbf{w}$ = div $(\nabla p) = \Delta p$ y $\mathbf{w} \cdot \mathbf{n} = \mathbf{n} \cdot \nabla p$. Usaremos esto para probar la existencia. Dado $\mathbf{w}$, tomaremos p como una solución del siguiente problema de Neumann

$$\Delta p = \text{div } \mathbf{w} \quad \text{en } D, \quad \text{con} \quad \frac{\partial p}{\partial n} = \mathbf{w} \cdot \mathbf{n} \quad \textit{en } \partial D.$$

Puesto que, en virtud del teorema de la divergencia

$$\int_D \text{div } \mathbf{w} \, dV = \int_{\partial D} \mathbf{w} \cdot \mathbf{n} \, dA,$$

sabemos que este problema tiene solución única salvo constante aditiva en p (ver *Nota 1.9*). Con esta elección de p, se define $\mathbf{u} = \mathbf{w} - \nabla p$. Luego, claramente se tienen las propiedades deseadas div $\mathbf{u} = 0$, y $\mathbf{u} \cdot \mathbf{n} = 0$ por la construcción de p. Para probar la unicidad, estableceremos primero la relación de ortogonalidad

$$\int_D \mathbf{u} \cdot \nabla p \, dV = 0.$$

En efecto, por la regla de la cadena

$$\operatorname{div}(p\mathbf{u}) = (\operatorname{div} \mathbf{u})p + \mathbf{u} \cdot \nabla p,$$

y por el teorema de la divergencia y div $\mathbf{u} = 0$, obtenemos

$$\int_D \mathbf{u} \cdot \nabla p \, dV = \int_D \operatorname{div} (\mathrm{p}\mathbf{u}) \, dV = \int_{\partial D} p\mathbf{u} \cdot \mathbf{n} \, dA = 0$$

pues $\mathbf{u} \cdot \mathbf{n} = 0$ en ∂D.
Supongamos que $\mathbf{w} = \mathbf{u}_1 + \nabla p_1 = \mathbf{u}_2 + \nabla p_2$. Entonces

$$0 = \mathbf{u}_1 - \mathbf{u}_2 + \nabla(p_1 - p_2).$$

Tomando el producto interior con $\mathbf{u}_1 - \mathbf{u}_2$ e integrando, obtenemos

$$0 = \int_D \left\{ \|\mathbf{u}_1 - \mathbf{u}_2\|^2 + (\mathbf{u}_1 - \mathbf{u}_2) \cdot \nabla(p_1 - p_2) \right\} \, dV = \int_D \|\mathbf{u}_1 - \mathbf{u}_2\|^2 \, dV$$

por la relación de ortogonalidad. Sigue que $\mathbf{u}_1 = \mathbf{u}_2$ y entonces, $\nabla p_1 = \nabla p_2$, lo que prueba la unicidad de la descomposición.

■

Nota 1.10 La ecuación $\Delta p = f, \partial p/\partial n = g$ tiene solución única (salvo constante aditiva en p) si y solo si $\int_D f \, dV = \int_{\partial D} g \, dA$ ([5]).

Es natural introducir un operador de proyección ortogonal $\mathbb{P}$ que nos permita desacoplar la presión de la velocidad. Este operador asigna a $\mathbf{w}$ su campo libre de divergencia $\mathbf{u}$. $\mathbb{P}$ está bien definido por el teorema anterior (*Teorema* 1.3). Nótese que por construcción $\mathbb{P}$ es un operador lineal y

$$\mathbf{w} = \mathbb{P}\mathbf{w} + \nabla p. \tag{1.12}$$

Véase también que

$$\mathbb{P}\mathbf{u} = \mathbf{u}$$

siempre que div $\mathbf{u} = 0$ y $\mathbf{u} \cdot \mathbf{n} = 0$, y que

$$\mathbb{P}(\nabla p) = 0.$$

Ahora trasladaremos estas ideas a las ecuaciones de Navier-Stokes para fluidos incompresibles homogéneos (1.9). Si aplicamos el operador $\mathbb{P}$ a ambos lados, obtenemos

$$\mathbb{P}(\partial_t \mathbf{u} + \nabla p) = \mathbb{P}\left(-(\mathbf{u}\cdot\nabla)\mathbf{u} + \frac{1}{R}\Delta\mathbf{u}\right).$$

Ya que $\mathbf{u}$ es libre de divergencia y se anula en la frontera, se cumple lo mismo para $\partial_t\mathbf{u}$ (si $\mathbf{u}$ es suficientemente regular). Luego, por (1.12), $\mathbb{P}\partial_t\mathbf{u} = \partial_t\mathbf{u}$, y como $\mathbb{P}(\nabla p) = 0$, obtenemos

$$\partial_t \mathbf{u} = \mathbb{P}\left(-(\mathbf{u}\cdot\nabla)\mathbf{u} + \frac{1}{R}\Delta\mathbf{u}\right). \tag{1.13}$$

Esta forma (1.13) de las ecuaciones de Navier-Stokes permite eliminar la presión y expresar $\partial_t\mathbf{u}$ únicamente en términos de $\mathbf{u}$. La presión se puede recuperar como la parte gradiente de

$$-(\mathbf{u}\cdot\nabla)\mathbf{u} + \frac{1}{R}\Delta\mathbf{u}.$$

Esto no tiene solo interés teórico sino que también es de interés práctico para algoritmos numéricos, como los que estudiaremos en el capítulo 3.

Nota 1.11 Si R es suficientemente pequeño, las ecuaciones (1.10) son aproximadas por

$$\partial_t \mathbf{u} = \mathbb{P}\left(\frac{1}{R}\Delta\mathbf{u}\right),$$

esto es,

$$\partial_t \mathbf{u} = -\nabla p + \frac{1}{R}\Delta\mathbf{u} \qquad \text{y} \qquad \operatorname{div}\mathbf{u} = 0,$$

lo que se conoce como las ecuaciones de Stokes para fluidos incompresibles. Son ecuaciones lineales parabólicas que, para R muy pequeño, proporcionan una buena aproximación de las ecuaciones de Navier-Stokes.

Finalizamos este capítulo inicial señalando que existe una diferencia importante entre un fluido ideal y uno viscoso en lo que respecta a la energía del mismo. En particular, para fluidos incompresibles, podemos introducir el siguiente resultado (ver *Nota 1.4*).

Lema 1.3 Para un fluido viscoso incompresible al que no se le aplican fuerzas externas, se tiene que

$$\frac{d}{dt}E_{\text{cinética}} \leq 0.$$

Demostración. Calculamos $\frac{d}{dt}E_{\text{cinética}}$ usando el teorema de transporte como hicimos en §**1.1**. Obtenemos

$$\begin{aligned}\frac{d}{dt}E_{\text{cinética}} &= \frac{d}{dt}\frac{1}{2}\int_D \rho\,\|\mathbf{u}\|^2\ dV = \int_D \rho\mathbf{u}\cdot\frac{D\mathbf{u}}{Dt}\ dV \\ &= \int_D \left(-\mathbf{u}\cdot\nabla p + \frac{1}{R}\mathbf{u}\cdot\Delta\mathbf{u}\right)\ dV,\end{aligned}$$

por (1.7) y div $\mathbf{u} = 0$. Como $\mathbf{u}$ es ortogonal a ∇p, obtenemos

$$\frac{d}{dt}E_{\text{cinética}} = \frac{1}{R}\int_D \mathbf{u}\cdot\Delta\mathbf{u}\ dV.$$

La relación $\operatorname{div}(f\mathbf{V}) = f\ \operatorname{div}\ \mathbf{V} + \mathbf{V}\cdot\nabla f$ nos lleva a

$$\begin{aligned}\operatorname{div}(\mathbf{u}\cdot\nabla\mathbf{u}) &= \nabla\cdot(u\nabla u + v\nabla v + w\nabla w)\\ &= u\Delta u + v\Delta v + w\Delta w + \nabla u\cdot\nabla u + \nabla v\cdot\nabla v + \nabla w\cdot\nabla w\\ &= \mathbf{u}\cdot\Delta\mathbf{u} + \|\nabla\mathbf{u}\|^2.\end{aligned}$$

Luego

$$\int_D \operatorname{div}(\mathbf{u}\cdot\nabla\mathbf{u})\ dV = \int_D \mathbf{u}\cdot\Delta\mathbf{u}\ dV + \int_D \|\nabla\mathbf{u}\|^2\ dV$$

y por el teorema de la divergencia y la condición $\mathbf{u} = 0$ en ∂D

$$\int_{\partial D}(\mathbf{u}\cdot\nabla\mathbf{u})\cdot\mathbf{n}\ dA = 0 \Rightarrow \int_D \mathbf{u}\cdot\Delta\mathbf{u}\ dV = -\int_D \|\nabla\mathbf{u}\|^2\ dV.$$

Finalmente llegamos a que

$$\frac{d}{dt}E_{\text{cinética}} = -\frac{1}{R}\int_D \|\nabla\mathbf{u}\|^2\ dV \le 0.$$

■

2

Métodos de un paso. Métodos de tipo Runge-Kutta

Es necesario introducir los métodos numéricos de un paso y hablar de su consistencia, estabilidad y convergencia. Esto luego nos permitirá entender los métodos tipo Runge-Kutta y, sobretodo, los SDIRK(Simply Diagonal Implicit Runge Kutta), los cuales serán implementados en los métodos de proyección que usaremos para las ecuaciones de Navier-Stokes en el capítulo 3.

2.1. Métodos de un paso: consistencia, estabilidad y convergencia

Consideramos el problema de valor inicial (PVI) en ecuaciones diferenciales ordinarias (EDO) de primer orden

$$\begin{cases} y' = f(t,y), t \in [t_0,T], y, f \in D \subseteq \mathbb{R}^m \\ y(t_0) = y_0 \end{cases} \tag{2.1}$$

donde $D \subseteq \mathbb{R}^m$ es abierto y convexo. Para garantizar la existencia y unicidad de la solución $y(t)$, definida al menos en un entorno del punto inicial t_0 asumiremos que $f \in C_L([t_0,T] \times \bar{D})$, esto es, f es continuamente Lipschitz en $[t_0,T] \times \bar{D}$, donde L es una constante de Lipschitz de f (respecto de la variable y). En lo que sigue asumiremos que $\|\cdot\|$ es una norma arbitraria prefijada en $\mathbb{R}^m$.

Tomando una partición de $[t_0,T]$, $t_0 < t_1 < \ldots < t_n < t_{n+1} < \ldots < t_N = T$ con $h_n := t_{n+1} - t_n, 0 \leq n \leq N-1$, tenemos que si $y(t)$ es solución de (2.1) en $[t_0,T]$, entonces:

$$\begin{aligned} y(t_{n+1}) = y(t_n + h_n) &= y(t_n) + h_n y'(t_n) + o(h_n) \\ &= y(t_n) + h_n f(t_n, y(t_n)) + o(h_n),\ h_n \to 0. \end{aligned}$$

El método de Euler explícito queda definido por la recurrencia

$$y_{n+1} = y_n + h_n \cdot f(t_n, y_n),\ n = 0, 1, \ldots, N-1.$$

El error de discretización local en $t = t_n$ es el error que comete el método tras dar un paso de tamaño t_n al partir de la solución exacta $y_n = y(t_n)$. Por ejemplo, para el método de Euler,

$$l(t_n, h_n) := y(t_n + h_n) - [y(t_n) + h_n \cdot f(t_n, y(t_n))]$$

Si asumimos $y \in C^2([t_0, T])$, el error local se puede acotar como sigue

$$\|l(t,h)\| \leq \frac{Y_2}{2} \cdot h^2, \text{con } Y_2 := \max_{t\in[t_0,T]} \|y''(t)\|$$

Se denominan errores globales del método a las cantidades $\varepsilon_n = \|y(t_n) - y_n\|, 0 \leq n \leq N$.

El método de Euler explícito forma parte de una familia de métodos numéricos que computan una aproximación y_{n+1} a la solución en t_{n+1}=$t_n + h_n$ utilizando la aproximación y_n en el punto $t = t_n$, donde h_n es el tamaño de paso. Estos métodos se denominan métodos de un paso.

Ejemplo 2.1

1. Método de Euler implícito:

$$y_{n+1} = y_n + h_n \cdot f(t_{n+1}, y_{n+1}),\ h_n = t_{n+1} - t_n$$

2. θ-métodos:

$$y_{n+1} = y_n + (1-\theta) \cdot h_n \cdot f(t_n, y_n) + \theta \cdot h_n \cdot f(t_{n+1}, y_{n+1}), \text{con } \theta \in [0,1]$$

3. Regla trapezoidal explícita:

$$y_{n+1} = y_n + \frac{h_n}{2} \cdot f(t_n, y_n) + \frac{h_n}{2} \cdot h_n \cdot f(t_{n+1}, y_n + h_n \cdot f(t_n, y_n))$$

4. Regla trapezoidal implícita (θ-método con $\theta = \frac{1}{2}$):

$$y_{n+1} = y_n + \frac{h_n}{2} \cdot [f(t_n, y_n) + f(t_{n+1}, y_{n+1})]$$

5. Regla explícita del punto medio (método de Runge de orden dos):

$$y_{n+1} = y_n + h_n \cdot f(t_n + \frac{h_n}{2}, y_n + \frac{h_n}{2} \cdot f(t_n, y_n))$$

6. Regla implícita del punto medio:

$$y_{n+1} = y_n + h_n \cdot f(t_n + \frac{h_n}{2}, \frac{y_n + y_{n+1}}{2})$$

Se puede formular un método de un paso usando la notación de Henrici

$$y_{n+1} = y_n + h_n \cdot \phi(t_n, y_n, h_n)$$

donde $\phi(t_n, y_n, h_n)$ se denomina función de incremento del método. Se dice que el método es implícito cuando ϕ está definida implícitamente por f y explícito en caso contrario.

Ejemplo 2.2

1. Euler explícito: $\phi(t, y, h) = f(t, y)$
2. Runge: $\phi(t, y, h) = f(t + \frac{h}{2}, y + \frac{h}{2} f(t, y))$
3. Euler implícito: $\phi(t, y, h) = f(t + h, y + h \cdot \phi)$

Definición 2.1 Dado el PVI (2.1), el operador de error local del método con paso h en el punto (t, z) es:

$$L[t, z, h] := y(t + h; t, z) - [z + h \cdot \phi(t, z, h)]$$

donde $y(t + h; t, z)$ es la solución exacta de la EDO con valor inicial $y(t) = z$ y ϕ la función incremento del método. Si z varía a lo largo de una curva integral, el error local del método viene dado por

$$l(t, h) := L[t, y(t), h] = y(t + h) - y(t) - h \cdot \phi(t, y(t), h)$$

Definición 2.2 Un método de un paso que verifique las hipótesis de existencia y unicidad del problema (2.1) se dice consistente si cumple $\lim_{h \to 0} \frac{\|l(t,h)\|}{h} = 0$ uniformemente en $t \in [t_0, T]$, para cualquier curva integral $y(t)$ de la EDO $y' = f(t, y)$. Es decir, cuando

$$\forall \varepsilon > 0, \exists \delta : 0 < |h| < \delta : \|l(t, h)\| \leq \varepsilon \cdot |h|, \forall t \in [t_0, T].$$

Si $f \in C^p([t_0, T] \times D)$, con $p \geq 1$, diremos que el método es consistente de orden p, si para toda curva integral $y(t)$ existen $K, \delta > 0$ tales que

$$\|l(t, h)\| \leq K \cdot |h|^{p+1}, \forall |h| \leq \delta, \forall t \in [t_0, T].$$

Teorema 2.1 (Caracterización de consistencia).
Sea $\phi(t, y, h) \in C([t_0, T] \times D \times (0, \bar{h}])$ la función de incremento de un método de un paso para el sistema $y' = f(t, y)$. El método es consistente si y solo si

$$\phi(t, y, 0) = f(t, y), \forall t, y.$$

Demostración. Teniendo en cuenta que

$$\frac{l(t, h)}{h} = \frac{y(t + h) - y(t) - h \cdot y'(t)}{h} - [\phi(t, y(t), h) - f(t, y(t))],$$

que ϕ es continua, y que $\lim_{h \to 0} \frac{y(t+h) - y(t) - h \cdot y'(t)}{h} = 0$, la demostración es inmediata.

■

Definición 2.3 Un método de un paso con función de incremento ϕ se dice estable si existen constantes $\delta, K > 0$ tales que para cualquier partición P de $[t_0, T]$ con diámetro $|P| \leq \delta$ se verifica que las siguientes secuencias

$$\begin{cases} y_{n+1} = y_n + h_n \cdot \phi(t_n, y_n, h_n), n \geq 0, \\ y_0 \text{ dado}, \end{cases} \tag{2.2}$$

$$\begin{cases} \bar{y}_{n+1} = \bar{y}_n + h_n \cdot \phi(t_n, \bar{y}_n, h_n) + \eta_{n+1}, n \geq 0, \\ \bar{y}_0 = y_0 + \eta_0, \end{cases} \quad 0 \leq n \leq N-1 \tag{2.3}$$

cumplen, para cualquier conjunto de perturbaciones $\{\eta_j\}_{j=0}^N$, que

$$\|\bar{y}_n - y_n\| \leq K \cdot \sum_{j=0}^{n} \|\eta_j\|, 0 \leq n \leq N.$$

Teorema 2.2. *Si la función incremento $\phi(t, y, h)$ de un método de un paso es lipschitziana respecto de y en $[t_0, T] \times D \times (0, \bar{h}]$ entonces el método es estable.*

Demostración. Sea L_ϕ una constante de Lipschitz de ϕ respecto de y. Entonces:

$$\|\bar{y}_{n+1} - y_{n+1}\| \leq \|\bar{y}_n - y_n\| \cdot (1 + h_n + L_\phi) + \|\eta_{n+1}\| \leq e^{L_\phi(t_{n+1} - t_n)} \cdot \|\bar{y}_n - y_n\| + \|\eta_{n+1}\|$$

Inductivamente:

$$\begin{aligned} \|\bar{y}_n - y_n\| &= \|\eta_n\| + e^{L_\phi(t_n - t_n - 1)} \cdot \|\bar{y}_{n-1} - y_{n-1}\| \\ &\leq \|\eta_n\| + \|\eta_{n-1}\| \cdot e^{L_\phi(t_n - t_{n-1})} + e^{L_\phi(t_n - t_{n-2})} \cdot \|\bar{y}_{n-2} - y_{n-2}\| \\ &\leq \|\eta_n\| + \|\eta_{n-1}\| \cdot e^{L_\phi(t_n - t_{n-1})} + \|\eta_{n-2}\| \cdot e^{L_\phi(t_n - t_{n-2})} + \cdots + \\ &+ \|\eta_1\| \cdot e^{L_\phi(t_n - t_1)} + \|\eta_0\| \cdot e^{L_\phi(t_n - t_0)} \\ &\leq e^{L_\phi(T - t_0)} \cdot \sum_{j=0}^{n} \|\eta_j\|, \ 0 \leq n \leq N \end{aligned}$$

Tomando $K = e^{L_\phi(T - t_0)}$, llegamos a que el método es estable.

■

Definición 2.4 Un método de un paso se dice convergente si $\forall \varepsilon > 0, \exists \delta > 0$ tal que para toda partición P de $[t_0, T], t_0 < t_1 < \ldots < t_N = T$, con diámetro $|P| < \delta$ se tiene que

$$\max_{t_n \in P} \|y(t_n) - y_n\| < \varepsilon,$$

en otras palabras, un método de un paso será convergente cuando

$$\lim_{|P| \to 0} \max_{0 \leq n \leq N} \|y(t_n) - y_n\| = 0$$

Definición 2.5 Sea el PVI (2.1) con $f \in C^p([t_0, T] \times D)$, para un cierto $p \geq 1$. Se dice que un método de un paso es convergente de orden p si existen $K, \delta > 0$, tales que para toda partición P de $[t_0, T]$ con $|P| < \delta$ se tiene que

$$\max_{t_n \in P} \|y(t_n) - y_n\| \leq K \cdot h^p, \quad \text{con } h = |P|.$$

A continuación, establecemos una relación entre convergencia, consistencia y estabilidad para métodos de un paso.

Teorema 2.3. *Sea ϕ la función incremento de un método de un paso tal que $\phi(t, y, h)$ es continua en $[t_0, T] \times D \times [0, \bar{h}]$ y lipschitziana respecto de y.*

1. *Si el método es consistente entonces es convergente.*
2. *Si además $f \in C^p([t_0, T] \times D)$ y el método es consistente de orden p, entonces es convergente de orden p.*

Demostración. 1. Sea $\varepsilon > 0$, por consistencia del método, existe $\delta > 0$ tal que para $|h| < \delta, \|l(t, h)\| \leq \varepsilon \cdot |h|$. Sea P una partición de $[t_0, T]$ con $|P| < \delta$. Entonces

$$y(t_{n+1}) = y(t_n) + h_n \cdot \phi(t_n, y(t_n), h_n) + l_n, 0 \leq n \leq N - 1,$$

con $l_n = l(t_n, h_n)$. Denotemos $d_n := \|y(t_n) - y_n\|$, donde $\{y_n\}_{n=0}^N$ son las soluciones numéricas dadas por el método:

$$y_{n+1} = y_n + h_n \cdot \phi(t_n, y, h_n), 0 \leq n \leq N - 1.$$

Puesto que ϕ es lipschitziana respecto de y:

$$d_{n+1} = \|y(t_{n+1} - y_{n+1})\| \leq (1 + h_n \cdot L_\phi) \|y(t_n) - y_n\| + \|l_n\|, 0 \leq n \leq N - 1.$$

Por tanto

$$d_n \leq e^{L_\phi(t_n - t_{n-1})} \cdot d_{n-1} + \|l_{n-1}\| \leq e^{L_\phi(t_n - t_0)} \cdot \sum_{j=0}^{n-1} \|l_j\| \leq e^{L_\phi(T - t_0)} \cdot \sum_{j=0}^{N-1} \|l_j\|$$

para cada $0 \leq n \leq N$. Finalmente, como $\|l_j\| \leq \varepsilon \cdot h_j, 0 \leq j \leq N - 1$,

$$d_n \leq e^{L_\phi(T - t_0)} \cdot \varepsilon \sum_{j=0}^{N-1} h_j = (T - t_0) \cdot e^{L_\phi(T - t_0)} \cdot \varepsilon \ = K \cdot \varepsilon.$$

Esto implica que el método es convergente.
2. La prueba de este apartado es análoga a la realizada en 1.

■

2.2. Métodos de tipo Runge-Kutta

A continuación, se presentan los métodos de tipo Runge-Kutta (RK), que constituyen una familia importante de métodos de un paso. Un método RK de s etapas aplicado al PVI (2.1) para avanzar con paso $h > 0$ desde $t = t_0$ hasta $t_1 = t_0 + h$ computando $y_1 \simeq y(t_0 + h)$ se define de la siguiente manera:

$$\begin{cases} K_i = f(t_0 + c_i \cdot h, y_0 + h \cdot \sum_{j=1}^{s} a_{ij} K_j), \ 1 \le i \le s, \\ y_1 = y_0 + h \cdot \sum_{i=1}^{s} b_i K_i, \end{cases} \tag{2.4}$$

donde $K_1, \ldots, K_s \in \mathbb{R}^m$ se denominan etapas internas del método. Los coeficientes $\{c_i\}_{i=1}^s, \{b_i\}_{i=1}^s, \{a_{ij}\}_{i,j=1}^s$ definen el método RK. De manera compacta, el método RK se puede representar mediante la tabla:

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \cdots & a_{1s} \\ c_2 & a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_s & a_{s1} & a_{s2} & \cdots & a_{ss} \\ \hline & b_1 & b_2 & \cdots & b_s \end{array} \equiv \begin{array}{c|c} c & A \\ \hline & b^T \end{array}$$

que se denomina tabla de Butcher del método, con $c := (c_1, \cdots, c_s)^T, A := (a_{ij})_{i,j=1}^s$ y $b := (b_1, \cdots, b_s)^T$. Denotaremos $e := (1, \cdots, 1)^T \in \mathbb{R}^s$.

Nota 2.1 Observamos que, en general, la ecuación de etapas de un método RK es un sistema implícito no lineal de dimensión $s \cdot m$. Sin embargo, si A es triangular inferior estricta la resolución de la ecuación de etapas resulta inmediata, puesto que cada etapa K_i se obtiene explícitamente a partir de las anteriores, y $K_1 = f(t_0 + c_1 \cdot h, y_0)$. En este caso, el método es explícito.

Nota 2.2 Todos los métodos del *Ejemplo 2.1* son métodos RK.

Podemos expresar el método $RK(A, b, c)$ utilizando las siguientes formulaciones, equivalentes a (2.4):

$$\begin{cases} Y_i = y_0 + h \cdot \sum_{j=1}^{s} a_{ij} f(t_0 + c_j h, Y_j), \ 1 \le i \le s, \\ y_1 = y_0 + h \cdot \sum_{i=1}^{s} b_i f(t_0 + c_i h, Y_i), \end{cases} \tag{2.5}$$

$$\begin{cases} g_i = h \cdot f(t_0 + c_i h, y_0 + \sum_{j=1}^{s} a_{ij} \cdot g_j), \ 1 \le i \le s, \\ y_1 = y_0 + \sum_{i=1}^{s} b_i g_i, \end{cases} \tag{2.6}$$

Ahora introducimos un resultado con el que aseguramos que la ecuación de etapas (2.4) admite una única solución $K_i = K_i(h), 1 \le i \le s$, para h suficientemente pequeño.

Teorema 2.4 (Existencia y unicidad de solución para la ecuación de etapas de un método RK).
Sea $f : [t_0, T] \times D \to \mathbb{R}^m$ continua y lipschitz respecto de y, con constante de lipschitz L, y $h_0 = (L \cdot \|A\|_\infty)^{-1}$ donde A es la matriz de coeficientes del método RK. Entonces, la ecuación de etapas (2.4) admite una solución única:

$$K_i = K_i(h), \quad 1 \leq i \leq s, \quad \textit{para } |h| < h_0.$$

Además, si $f \in C^p([t_0, T] \times D), p \geq 1$ entonces, $K_i = K_i(h) \in C^p(-h_0, h_0), 1 \leq i \leq s$.

Nota 2.3 Recordemos que $\|A\|_\infty = \sup_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty} = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|$, para $A \in \mathbb{R}^{m \times n}$

Demostración. Definimos los supervectores $K = (K_1^T, \ldots, K_s^T)^T \in \mathbb{R}^{s \cdot m}$ y $F(h, K) = (f(t_0 + c_1 h, y_0 + h \cdot \sum_{j=1}^{s} a_{1j} K_j)^T, \cdots, f(t_0 + c_s h, y_0 + h \cdot \sum_{j=1}^{s} a_{sj} K_j)^T)^T \in \mathbb{R}^{s \cdot m}$. La ecuación de etapas (2.4) equivale a resolver la ecuación implícita

$$K = F(h, K).$$

Como consecuencia del Teorema del Punto Fijo, bastará comprobar que F es contractiva (respecto a K) para $|h| < h_0$. Considerando la norma en $\mathbb{R}^{s \cdot m}$: $|||V||| := \max_{1 \leq i \leq s} \|V_i\|$, $V = (V_1^T, \ldots, V_s^T)^T$, $V_i \in \mathbb{R}^m, 1 \leq i \leq s$, tenemos que

$$\begin{aligned} |||F(h, K) - F(h, \widetilde{K})||| &\leq L \cdot |h| \cdot \max_{1 \leq i \leq s} \| \sum_{j=1}^{s} a_{ij} (K_j - \widetilde{K}) \| \\ &\leq L \cdot |h| \cdot (\max_{1 \leq i \leq s} \sum_{j=1}^{s} |a_{ij}|) \ \|(K_j - \widetilde{K}_j)\| \end{aligned}$$

siendo la constante de contracción $L \cdot |h| \cdot \|A\|_\infty < 1$ por hipótesis ($|h| < h_0$). En definitiva, el teorema del Punto Fijo asegura la existencia y unicidad de solución para $K = K(h)$, si $|h| < h_0$. Además, teniendo en cuenta el teorema de la función implícita (en $\mathbb{R}^{m \cdot s}$) y la función $G(h, K) = K - F(h, K)$, vemos que

$$\frac{\partial G_i}{\partial K_j} = \frac{\partial K_i}{\partial K_j} - h a_{ij} \cdot \frac{\partial f}{\partial y}(t_0 + c_i \cdot h, y_0 + h \cdot \sum_{j=1}^{s} a_{ij} K_j),$$

y por tanto $\frac{\partial G_i}{\partial K_j}(h = 0, K) = \begin{cases} I_m, i = j \\ 0, i \neq j \end{cases}, 1 \leq i, j \leq s$, siendo I_m la matriz identidad de orden m. Así $\frac{\partial G}{\partial K}(h = 0, K) = I_{sm}$, y, en particular, $|\frac{\partial G}{\partial K}(h = 0, K)| \neq 0$. El teorema de la Función Implícita permite asegurar entonces que la solución de $K = F(h, K)$ tiene por los menos la misma regularidad que F, esto es,

$K = K(h) \in C^p(-h_0, h_0)$.

■

A continuación estudiamos la consistencia y estabilidad de los métodos Runge-Kutta, a su vez establecemos un resultado que caracteriza la convergencia de tales métodos.

Teorema 2.5. *Un método $RK(A, b, c)$ es consistente si y sólo si $b^T e = 1$.*

Demostración. Dado que la solución de avance del método $RK(A, b, c)$ es

$$y_1 = y_0 + h \cdot \sum_{i=1}^{s} b_i K_i,$$

la función incremento del método es $\phi(t, y, h) = \sum_{i=1}^{s} b_i K_i$, donde $K_i = K_i(t, y, h), 1 \leq i \leq s$, son funciones continuas en $[t_0, T] \times D \times [0, h_0]$ (ver *Teorema* 2.4). Además

$$\phi(t, y, 0) = \sum_{i=1}^{s} b_i K_i(t, y, 0) = \sum_{i=1}^{s} b_i f(t, y) = f(t, y) \Leftrightarrow \sum_{i=1}^{s} b_i = 1.$$

La consistencia es consecuencia del *Teorema* 2.1.

■

Teorema 2.6. *Sea $\phi(t, y, h) = \sum_{i=1}^{s} b_i K_i(t, y, h)$ la función incremento de un método $RK(A, b, c)$. Entonces ϕ es lipschitziana respecto de y en $[t_0, T] \times D \times [0, h_0]$, para $h_0 < (L \cdot \rho(|A|))^{-1}$, con constante de lipschitz*

$$L_\phi = L \cdot |b|^T \cdot (I - h_0 \cdot L \cdot |A|)^{-1} \cdot e,$$

siendo $|b| := (|b_1|, \ldots, |b_s|)^T, |A| := (|a_{ij}|)_{i,j=1}^{s}$, $\rho(\cdot)$ el radio espectral de una matriz y L una constante de lipschitz de $f(t, y)$ respecto de y.

Demostración.
Pongamos $\phi := \phi(t, y, h) = \sum_{i=1}^{s} b_i K_i$ y $\bar{\phi} := \phi(t, \bar{y}, h) = \sum_{i=1}^{s} b_i \bar{K}_i$, donde

$$K_i = f(t + c_i h, y + h \cdot \sum_{j=1}^{s} a_{ij} K_j), \quad \bar{K}_i = f(t + c_i h, \bar{y} + h \cdot \sum_{j=1}^{s} a_{ij} \bar{K}_j), \ 1 \leq i \leq s.$$

Dado que f es Lipschitz respecto de y:

$$\left\| \phi - \bar{\phi} \right\| \leq \sum_{i=1}^{s} |b_i| \cdot \left\| K_i - \bar{K}_i \right\|, \text{ con}$$

$$\left\|K_i - \bar{K}_i\right\| \leq L \cdot (\|y - \bar{y}\| + |h| \cdot \sum_{j=1}^{s} a_{ij} \left\|K_j - \bar{K}_j\right\|),\ 1 \leq i \leq s.$$

Definiendo $\Delta K := (\left\|K_1 - \bar{K}_1\right\|, \ldots, \left\|K_s - \bar{K}_s\right\|)^T$, tenemos que

$$\Delta K \leq L \cdot \|y - \bar{y}\|\, e + |h_0| \cdot L \cdot |A| \cdot \Delta K,$$

estos es $(I - h_0 \cdot L \cdot |A|)\Delta K \leq L \cdot \|y - \bar{y}\|\, e$, donde la desigualdad anterior se entiende componente a componente. Ahora bien, $\rho(h_0 \cdot L \cdot |A|) = h_0 \cdot L \cdot \rho(|A|) < 1$ por hipótesis, y por el Teorema de Neumann

$$I - h_0 \cdot L \cdot |A|, \text{ es inversible y } [I - h_0 \cdot L \cdot |A|]^{-1} = \sum_{l=0}^{\infty} (h_0 \cdot L \cdot |A|)^l.$$

En particular, $[I - h_0 \cdot L \cdot |A|]^{-1}$ tiene todas sus componentes no negativas. Por lo tanto:

$$\Delta K \leq L \cdot \|y - \bar{y}\| \cdot (I - h_0 \cdot L \cdot |A|)^{-1} \cdot e.$$

Finalmente tenemos que

$$\left\|\phi - \bar{\phi}\right\| \leq |b|^T \cdot \Delta K \leq (L \cdot |b|^T (I - h_0 \cdot L \cdot |A|)^{-1} \cdot e) \cdot \|y - \bar{y}\|.$$

■

Corolario 2.1 Si un método $RK(A, b, c)$ cumple $b^T e = 1$ entonces es convergente.

Nota 2.4 La condición $b^T e = 1$ asegura convergencia de orden 1. La condición adicional $b^T c = \frac{1}{2}$ asegura orden 2. Si además se cumplen $b^T c^2 = \frac{1}{3}$ y $b^T A c = \frac{1}{6}$ entonces el método es de orden 3 (véase [2],[3],[6]).

2.3. Resolución de la ecuación de etapas en métodos Runge-Kutta implícitos. Métodos DIRK.

Sea la ecuación de etapas (2.6) de un método RK implícito de s etapas:

$$K_i = h \cdot f(t_n + c_i \cdot h, y_n + \sum_{j=1}^{s} a_{ij} \cdot K_j), 1 \leq i \leq s,$$

consideremos $K = (K_1^T, \ldots, K_s^T)^T \in \mathbb{R}^{m \cdot s}$ y

$$F(t_n \cdot e + h \cdot c, e \otimes y_n + (A \otimes I)K) := (f(t_n + c_i \cdot h, y_n + \sum_{j=1}^{s} a_{ij} K_j)^T)_{i=1,\ldots,s}^T \in \mathbb{R}^{m \cdot s}.$$

En forma compacta, usando el producto de Kronecker $A \otimes B = (a_{ij}B)_{i,j=1}^{s}$ la ecuación de etapas queda como:

$$K = h \cdot F(t_n \cdot e + h \cdot c, e \otimes y_n + (A \otimes I)K), e = (1, \ldots, 1)^T \in \mathbb{R}^{m \cdot s}. \tag{2.7}$$

donde $c = (c_1, \ldots, c_s)^T$ y

$$e \otimes y_n + (A \otimes I)K = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \otimes y_n + \begin{pmatrix} a_{11}I_n & \cdots & a_{1s}I_n \\ \vdots & & \vdots \\ a_{s1}I_n & \cdots & a_{ss}I_n \end{pmatrix} \cdot \begin{pmatrix} K_1 \\ \vdots \\ K_s \end{pmatrix} = \begin{pmatrix} y_n + \sum_{j=1}^{s} a_{1j} \cdot K_j \\ \vdots \\ y_n + \sum_{j=1}^{s} a_{sj} \cdot K_j \end{pmatrix}$$

En general el cálculo efectivo de la solución (2.7) se lleva a cabo mediante iteraciones de tipo Newton. Definimos

$$G(K) := K - h \cdot F(t_n \cdot e + hc, e \otimes y_n + (A \otimes I)K).$$

El método de Newton sobre el sistema $G(K) = 0$ será

$$G'(K^{(\nu)}) \cdot (K^{(\nu+1)} - K^{(\nu)}) = -G(K^{(\nu)}),\ \nu \geq 0,$$

o bien, introduciendo $\Delta^{(\nu)} := K^{(\nu+1)} - K^{(\nu)}$:

$$\begin{cases} G'(K^{(\nu)}) \cdot \Delta^{(\nu)} = -G(K^{(\nu)}) \\ K^{(\nu+1)} := K^{(\nu)} + \Delta^{(\nu)} \end{cases}, \quad \nu \geq 0. \tag{2.8}$$

Una elección natural posible para $K^{(0)}$ es $K_i^{(0)} = 0, 1 \leq i \leq s$. Observamos que $G(K) = (G_1^T, \ldots, G_s^T)^T$ con $G_i = K_i - hf(t_n + c_i h, y_n + \sum_{j=1}^{s} a_{ij}K_j), 1 \leq i \leq s$ y entonces $\frac{\partial G_i}{\partial K_j} = \delta_{ij} \cdot I_m - h \cdot a_{ij} \cdot J_j, 1 \leq i, j \leq s$, siendo δ_{ij} la delta de Kronecker y $J_i = \frac{\partial f}{\partial y}(t_n + c_i \cdot h, y_n + \sum_{j=1}^{s} a_{ij}K_j), 1 \leq i \leq s$. Luego, la matriz jacobiana $G'(K)$ viene dada por

$$G'(K) = \begin{bmatrix} I_m - ha_{11} \cdot J_1 & -ha_{12} \cdot J_2 & \cdots & -ha_{1s} \cdot J_s \\ -ha_{21} \cdot J_1 & I_m - ha_{22} \cdot J_2 & \cdots & -ha_{2s} \cdot J_s \\ \vdots & \vdots & \ddots & \vdots \\ -ha_{s1} \cdot J_1 & -ha_{1s2} \cdot J_2 & \cdots & I_m - ha_{ss} \cdot J_s \end{bmatrix} \in \mathbb{R}^{(sm)\times(sm)}. \tag{2.9}$$

A efectos de reducir el costo computacional involucrado en (2.8)-(2.9), una posibilidad consiste en reemplazar J_i por $J := \frac{\partial f}{\partial y}(t_n, y_n), 1 \leq i \leq s$, lo cual da lugar al método de Newton simplificado:

$$\begin{cases} [I_{ms} - h(A \otimes J)] \cdot \Delta^{(\nu)} = -G(K^{(\nu)}) \\ K^{(\nu+1)} := K^{(\nu)} + \Delta^{(\nu)} \end{cases}, \quad \nu \geq 0. \tag{2.10}$$

La implementación del esquema iterativo (2.10) requiere una descomposición LU de la matriz $I_{ms} - h(A \otimes J) \in \mathbb{R}^{(sm)\times(sm)}$; esto puede ser bastante costoso si la

dimensión m del PVI es grande. El esquema (2.10) es simple de implementar en el caso de métodos diagonalmente implícitos (DIRK), para los que $a_{ij} = 0$, si $j > i$. En tal caso, en vez de una factorización LU para la matriz de dimensión ms $I_{ms} - h(A \otimes J)$, se reduce el álgebra a s factorizaciones LU para las matrices de dimensión m $I_m - ha_{ii}J, 1 \leq i \leq s$. Aun más convenientes son los métodos simplemente diagonalmente implícitos (SDIRK) para los que $a_{ij} = 0, j > i$, y $a_{ii} = \gamma, 1 \leq i \leq s$, que solo requieren una descomposición LU para la matriz de dimensión m $I_m - h\gamma J$ en cada paso de la integración temporal:

$$\begin{bmatrix} I - h\gamma J & 0 & \cdots & 0 \\ -ha_{21}J & I - h\gamma J & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -ha_{s1}J & -ha_{s2}J & \cdots & I - h\gamma J \end{bmatrix} \cdot \begin{bmatrix} \Delta_1^{(\nu)} \\ \Delta_2^{(\nu)} \\ \vdots \\ \Delta_s^{(\nu)} \end{bmatrix} = - \begin{bmatrix} G_1^{(\nu)} \\ G_2^{(\nu)} \\ \vdots \\ G_s^{(\nu)} \end{bmatrix} \quad , \ \nu \geq 0, \qquad (2.11)$$

o bien, para $\nu \geq 0$,

$$\begin{cases} (I - h\gamma J)\Delta_i^{(\nu)} = -G_i^{(\nu)} + h \cdot \sum_{j=1}^{i-1} a_{ij} J \Delta_j^{(\nu)}, & 1 \leq i \leq s, \\ K_i^{(\nu+1)} = K_i^{(\nu)} + \Delta_i^{(\nu)} \end{cases}$$

siendo $G_i^{(\nu)} = K_i^{(\nu)} - h \cdot f(t_n + c_i \cdot h, y_n + \sum_{j=1}^{i} a_{ij} K_j^{(\nu)}),\ 1 \leq i \leq s$.

Teorema 2.7. *El orden de consistencia p de un método DIRK(A,b) de s etapas verifica* $p \leq s + 1$.

Demostración. Sea el método DIRK(A,b) aplicado al problema escalar autónomo

$$\begin{cases} y' = \lambda y \\ y(0) = y_0 \end{cases}, \lambda \in \mathbb{C},$$

cuya solución exacta es $y(t) = e^{\lambda t} \cdot y_0, t \geq 0$.

$$\begin{cases} K_i = h \cdot \lambda[y_0 + \sum_{j=1}^{i} a_{ij} K_j],\ 1 \leq i \leq s, \\ y_1 = y_0 + \sum_{i=1}^{s} b_i K_i \end{cases}$$

Poniendo $z := \lambda h$ y $K = (K_1, \ldots, K_s)^T \in \mathbb{C}^s$. Tenemos que,

$$\begin{cases} K = zy_0 e + zAK \\ y_1 = y_0 + b^T K \end{cases}, \text{o bien } \begin{bmatrix} I - zA & 0 \\ -zb^T & 1 \end{bmatrix} \cdot \begin{bmatrix} K \\ z \cdot y_1 \end{bmatrix} = \begin{bmatrix} e \\ 1 \end{bmatrix} zy_0.$$

Aplicando la regla de Cramer para obtener $z \cdot y_1$:

$$z \cdot y_1 = \frac{\det \begin{bmatrix} I - zA & e \cdot (zy_0) \\ -zb^T & 1 \cdot (zy_0) \end{bmatrix}}{\det \begin{bmatrix} I - zA & 0 \\ -zb^T & 1 \end{bmatrix}} = zy_0 \cdot \frac{\det \begin{bmatrix} I - zA & e \\ -zb^T & 1 \end{bmatrix}}{\det(I - zA)} = zy_0 \cdot \frac{\det \begin{bmatrix} I - zA + zeb^T & 0 \\ -zb^T & 1 \end{bmatrix}}{\det(I - zA)}$$

Luego $y_1 = y_0 \cdot \dfrac{\det(I - zA + zeb^T)}{\det(I - zA)}$. Puesto que $A \in \mathbb{R}^{s\times s}$ es triangular inferior, $R(z) := \dfrac{\det(I - zA + zeb^T)}{\det(I - zA)} = \dfrac{P_s(z)}{\prod_{i=1}^{s}(1 - za_{ii})}$ es una función racional con numerador y denominador de grado menor o igual que s, con s polos reales $(\frac{1}{a_{ii}}, 1 \leq i \leq s)$. Usando [7, Teorema 4.18, pág. 61] se tiene que $e^z - R(z) = \mathcal{O}(z^{p+1})$, con $p \leq s + 1$. Así, $y(h) - y_1 = y_0 \cdot [e^z - R(z)] = \mathcal{O}(h^{p+1}),\ h \to 0$, con $p \leq s + 1$.

■

2.4. Estabilidad absoluta lineal de los métodos Runge-Kutta

El concepto de estabilidad conocido busca acotar los errores globales del método en función de los errores locales introducidos en cada paso para un intervalo finito. Así, dado el método

$$y_{n+1} = y_n + h \cdot \phi(t_n, y_n, h),\ 0 \leq n \leq N - 1,\ N = \frac{T - t_0}{h} \tag{2.12}$$

y el método perturbado

$$\begin{cases} \tilde{y}_{n+1} = \tilde{y}_n + h \cdot \phi(t_n, \tilde{y}_n, h) + h \cdot \delta_{n+1},\ 0 \leq n \leq N - 1,\ N = \dfrac{T - t_0}{h} \\ \tilde{y}_0 = y_0 + \delta_0, \end{cases} \tag{2.13}$$

el método es estable si existen constantes k y h_0 tales que

$$\|\tilde{y}_n - y_n\| \leq k \cdot \max_{0\leq k\leq N} \left\| \delta_0 + \sum_{j=1}^{k} h \cdot \delta_j \right\|,\ 0 \leq n \leq N, \tag{2.14}$$

para todo $0 < h \leq h_0$ con $N \cdot h = T - t_0$.

Los errores globales se pueden acotar en función de los errores locales $l_n = y(t_n + h) - y(t_n) - h \cdot \phi(t_n, y(t_n), h)$ a lo largo de la solución de la forma

$$\|y(t_n) - y_n\| \leq K \cdot \max_{0\leq k\leq N-1} \left\| \sum_{j=0}^{k} l_j \right\|,\ 0 \leq n \leq N \tag{2.15}$$

Nótese que las ecuaciones (2.12) y (2.13) son, respectivamente, discretizaciones del PVI

$$\begin{cases} y'(t) = f(t, y(t)),\ t \in [t_0, T] \\ y(t_0) = y_0 \end{cases} \tag{2.16}$$

y del problema perturbado

$$\begin{cases} \tilde{y}'(t) = f(t, \tilde{y}(t)) + \delta(t), \ t \in [t_0, T] \\ \tilde{y}(t_0) = y_0 + \delta_0 \end{cases} \tag{2.17}$$

Si $f(t, y)$ verifica la condición de Lipschitz (respecto de y) se deduce a partir del lema de Gronwall una acotación de la forma

$$\|\tilde{y}(t) - y(t)\| \leq e^{\lambda(T-t_0)} \cdot \max_{t \in [t_0, T]} \left\{ \delta_0 + \int_{t_0}^{t} \delta(s) \ ds \right\} \tag{2.18}$$

que es la versión continua de (2.14) e implica que la solución de (2.16) depende continuamente de los datos (y_0, f) ((2.16) es totalmente estable).

A continuación estudiaremos un concepto de estabilidad que requiere cierta continuidad del método (2.12) respecto a los datos, pero en intervalos no acotados de la variable independiente (este concepto se va a corresponder con el concepto de estabilidad de Lyapunov).

Definición 1.5 Sea un PVI $\begin{cases} y'(t) = f(t, y(t)) \\ y(t_0) = y_0 \end{cases}$ cuya solución única $y(t)$ está definida en el intervalo $[t_0, +\infty)$ y $f \in C^1$ en un entorno $T_\nu = \{(t, y)/t \in [t_0, +\infty), |y - y(t)| \leq \nu\}$ de la solución considerada. La solución $y(t)$ se dice:

- Estable: si $\forall \varepsilon > 0, \ \forall t_1 \geq t_0, \ \exists \delta = \delta(\varepsilon, t_1)$ tal que $|\tilde{y}_1 - y(t_1)| < \delta \Rightarrow$ $\Rightarrow |y(t; t_1, \tilde{y}_1) - y(t)| < \varepsilon, \forall t \geq t_1$.
- Asintóticamente estable: si es estable y además existe $\gamma_1 = \gamma(t_1)$ tal que $\lim_{t \to \infty} |y(t; t_1, \tilde{y}_1) - y(t)| = 0$, para todo $\tilde{y}_1$ tal que $|\tilde{y}_1 - y(t_1)| < \gamma_1$.

Sería deseable que cuando un método numérico se aplica a un PVI estable (o asintóticamente estable) la solución numérica exhibiera el mismo comportamiento de estabilidad. Este requerimiento es difícil de satisfacer por los métodos numéricos por lo que se consideran familias particulares de ecuaciones diferenciales cuyas soluciones poseen propiedades de estabilidad bien conocidas y se estudian los métodos numéricos sobre dichas ecuaciones. Introducimos el concepto de contractividad que es en esencia una versión discreta del concepto de estabilidad para EDO.

Definición 2.6 Dado el método de un paso

$$y_{n+1} = y_n + h \cdot \phi(t_n, y_n, h), \ n \geq 0 \tag{2.19}$$

(a) Se dice que es contractivo en y_0 si existen $h_0 > 0, \ K \geq 1$ y V_0 un entorno de y_0 de modo que la solución generada a partir de cualquier $\tilde{y}_0 \in V_0$ verifica

$$\|\tilde{y}_{n+1} - y_{n+1}\| \leq K \cdot \|\tilde{y}_n - y_n\| \tag{2.20}$$

(b) Si se verifica lo anterior con $K < 1$, el esquema (2.19) se dice estrictamente contractivo en y_0.
(c) Si se verifica (2.20) para todo $h \geq 0$, el método (2.19) se dice incondicionalmente contractivo.

En adelante consideramos el caso sencillo de ecuaciones lineales de la forma:

$$y' = \lambda y \quad \text{con } \lambda \in \mathbb{C},\ Re\lambda \leq 0. \tag{2.21}$$

Se puede ver que toda solución de (2.21) es estable (respectivamente asintóticamente estable) si y sólo si $Re\lambda \leq 0$ (respectivamente $Re\lambda < 0$) pues las soluciones de (2.21) son de la forma

$$y(t) = y(t_0) \cdot e^{\lambda(t-t_0)}$$

Así aplicando el método (2.19) a (2.21), deberemos estudiar si el esquema obtenido es contractivo para $Re\lambda \leq 0$. Veremos posteriormente que cuando un método RK se aplica a la ecuación (2.21) con paso fijo h se obtiene que $y_{n+1} = R(h\cdot\lambda)\cdot y_n$, siendo R una función racional. En consecuencia, la condición de contractividad en cualquier punto se reduce a $|R(h \cdot \lambda)| \leq 1$. El estudio de estas propiedades se conoce como teoría de estabilidad absoluta lineal.

La propiedad de contractividad permite obtener acotaciones realistas para el error global. Supongamos que el método (2.19) verifica la condición de contractividad (2.20) y tomemos

$$\begin{cases} \hat{y}_{n+1} = \hat{y}_n + h \cdot \phi(t_n, \hat{y}_n, h) + w_{n+1},\ n \geq 0 \\ \hat{y}_0 = y_0 + w_0 \end{cases}$$

el esquema perturbado.

Aplicando (2.20) con $\tilde{y}_{n+1} = \hat{y}_{n+1} - w_{n+1}$ e $\tilde{y}_n = \hat{y}_n$ (a partir de $\hat{y}_0 = y_0 + w_0$ el método devolvería $\hat{y}_1 - w_1$) se tiene que

$$|\hat{y}_{n+1} - y_{n+1}| \leq K \cdot |\hat{y}_n - y_n| + |w_{n+1}|$$

y de aquí que

$$|\hat{y}_n - y_n| \leq \sum_{j=0}^{n} |w_j| \cdot K^{n-j}$$

En consecuencia, para los errores globales (con $w_j = l_j = y(t_j + h) - y(t_j) - h \cdot \phi(t_j, y(t_j), h)$) se tendrá que:

$$|y(t_n) - y_n| \leq \sum_{j=0}^{n} |l_j| \cdot K^{n-j} \leq \sum_{j=0}^{n} |l_j|,\ N \geq n \geq 0,$$

que garantiza un buen comportamiento de estabilidad del método en el sentido de que los errores globales quedan controlados por los errores locales en tiempos anteriores.

Problemas diferenciales stiff

En la práctica se presentan ciertos sistemas diferenciales llamados *stiff* para los cuales los métodos de integración explícitos resultan ineficientes. En estos problemas la derivada de f tiene por lo general constante de Lipschitz grande, que hace que la constante de estabilidad $e^{L(T-t_0)}$ sea grande, por lo que la acotación (2.15) para los errores globales no permite asegurar un buen comportamiento del método. Cuando la solución de un PVI varía lentamente en un intervalo y al mismo tiempo es extremadamente estable (las soluciones próximas tienden rápidamente a la original), el problema se dice *stiff* en este intervalo ([2],[3],[7]).

Consideremos el problema lineal

$$\begin{cases} y'(t) = \lambda \cdot (y(t) - p(t)) + p'(t), \ t > 0 \\ y(0) = y_0 \end{cases} \tag{2.22}$$

donde $\lambda \in \mathbb{C}$ y $p(t)$ es una función suficientemente diferenciable. La solución de esta ecuación es

$$y(t) = p(t) + (y_0 - p(0)) \cdot e^{\lambda t} \tag{2.23}$$

y por tanto para todo par de condiciones iniciales $y_0, \tilde{y}_0$ las soluciones correspondientes $y(t), \tilde{y}(t)$ verifican $\tilde{y}(t) - y(t) = (\tilde{y}_0 - y_0) \cdot e^{\lambda t}$
Vemos entonces que si $Re\lambda$ es positivo y grande, todo par de soluciones de (2.22) se separan rápidamente con el tiempo. Así, el problema es muy inestable y no cabe esperar que ningún método numérico que se aplique proporcione buenos resultados, ya que los errores cometidos crecerán rápidamente con el tiempo.
Afortunadamente estos problemas no son, en general, interesantes, ya que en muchos sistemas las variables de estado toman usualmente valores acotados.
Ahora bien, si $|\lambda|$ es pequeño las curvas solución en intervalos moderados de tiempo son aproximadamente paralelas. Este tipo de problemas se dice que tienen estabilidad neutra y se resuelven bien por medio de métodos explícitos.
Por último, si $Re\lambda$ es negativo y grande, todas las curvas solución de (2.22) tienden rápidamente a la solución estacionaria $p(t)$. A la segunda componente de (2.23) se le llama componente transitoria, ya que desaparece rápidamente con el tiempo.
Este sistema superestable muestra un comportamiento muy favorable para evitar la propagación de errores en la ecuación diferencial, pero sin embargo no ocurre lo mismo para la solución obtenida mediante RK explícitos.

A-estabilidad de los métodos Runge-Kutta

En la teoría de estabilidad absoluta se estudia el comportamiento de los métodos numéricos en intervalos no acotados de la variable independiente cuando se aplican a ecuaciones diferenciales lineales escalares de la forma

$$y'(t) = \lambda \cdot y(t) \quad \text{donde } \lambda \in \mathbb{C} \text{ y } t \geq 0. \tag{2.24}$$

Ya sabemos que las soluciones de (2.24) son estables en $t \geq 0$ si y solo si $Re\lambda \leq 0$, y son asintóticamente estables si y solo si $Re\lambda < 0$.

Al aplicar el $RK(A, b)$ de s etapas con paso h a la ecuación (2.22), con $y(0) = y_0$, se tiene

$$\begin{cases} Y_i = y_n + h \cdot \sum_{j=1}^{s} a_{ij} f(t_n + c_j h, Y_j), \ 1 \leq i \leq s \\ y_{n+1} = y_n + h \cdot \sum_{i=1}^{s} b_i f(t_n + c_i h, Y_i) \end{cases} \Leftrightarrow$$

$$\begin{cases} Y_i = y_n + h \cdot \lambda \cdot \sum_{j=1}^{s} a_{ij} \cdot Y_j, \ 1 \leq i \leq s \\ y_{n+1} = y_n + h \cdot \lambda \cdot \sum_{i=1}^{s} b_i \cdot Y_i \end{cases} \tag{2.25}$$

Llamando $Y = (Y_1, \ldots, Y_s) \in \mathbb{C}^s$, (2.25) puede reescribirse en la forma

$$\begin{cases} Y = y_n \cdot e + h \cdot \lambda \cdot A \cdot Y \\ y_{n+1} = y_n + h \cdot \lambda \cdot b^T \cdot Y \end{cases}$$

y si $det(I - h\lambda \cdot A) \neq 0$, sigue que

$$y_{n+1} = [1 + h \cdot \lambda \cdot b^T \cdot (I - h \cdot \lambda \cdot A)^{-1} \cdot e] \cdot y_n \tag{2.26}$$

(la matriz $I - h \cdot \lambda \cdot A$ es regular para valores de h suficientemente pequeños).

Definición 2.7 La función de variable compleja $R(z) = 1 + z \cdot b^T \cdot (I - z \cdot A)^{-1} \cdot e$ se llama función de estabilidad lineal del método RK(A, b).

Nota 2.5 La ecuación en diferencias resultante de aplicar el RK a la ecuación de prueba (2.24) será estable si y solo si $|R(z)| \leq 1$. Sería deseable que el método conservase las propiedades de estabilidad de la ecuación de prueba. Es decir, que para todo λ con $Re\lambda \leq 0$ y $h \geq 0$ se tuviera $|R(h \cdot \lambda)| \leq 1$, pero esto no se cumple en general. Por ejemplo, para el método de Euler:

$$R(h \cdot \lambda) = 1 + h \cdot \lambda \text{ y } |1 + h \cdot \lambda| \leq 1 \Leftrightarrow h \cdot \lambda \in \bar{D}(-1, 1)$$

siendo $\bar{D}(z_0, r)$ el disco cerrado de centro z_0 y radio r.

Así si se integra la ecuación $y' = -40y$ con paso $h = 0.1$, usando el método de Euler se tendrá:

$$y_{n+1} = y_n + h \cdot f(t_n, y_n) = (1-4)y_n = (-3)y_n = (-3)^{n+1} \cdot y_0$$

mientras que la solución analítica será $y(t_{n+1}) = y_0 \cdot e^{-40 \cdot t_{n+1}}$, por lo que para $y_0 \neq 0$ ambas soluciones muestran un comportamiento totalmente distinto.

Lema 2.1 $R(z) = \dfrac{det[I - z \cdot (A - e \cdot b^T)]}{det[I - z \cdot A]}$

Demostración. Tenemos

$$\begin{cases} Y = e \cdot y_n + z \cdot A \cdot Y \\ y_{n+1} = y_n + z \cdot b^T \cdot Y \end{cases}, (z = h \cdot \lambda) \text{ e } y_{n+1} = R(z) \cdot y_n$$

Sean $\hat{Y} = \begin{pmatrix} Y \\ y_{n+1} \end{pmatrix}, \hat{e} = (e^T, 1)^T$, luego $\hat{Y} = \hat{e} \cdot y_n + z \cdot \hat{A} \cdot \hat{Y}$, con $\hat{A} = \left(\begin{array}{c|c} A & \mathbf{0} \\ \hline b^T & 0 \end{array}\right)$

Así:

$$\begin{bmatrix} I - z \cdot A & 0 \\ -z \cdot b^T & 1 \end{bmatrix} \cdot \hat{Y} = \begin{pmatrix} e \\ 1 \end{pmatrix} \cdot y_n.$$

Por la regla de Cramer, si $det(I - z \cdot A) \neq 0$,

$$y_{n+1} = \frac{det \begin{bmatrix} I - z \cdot A & e \cdot y_n \\ -z \cdot b^T & y_n \end{bmatrix}}{det(I - z \cdot A)} = \frac{det \begin{bmatrix} I - z \cdot A & e \\ -z \cdot b^T & 1 \end{bmatrix}}{det(I - z \cdot A)} \cdot y_n = \frac{det[I - z \cdot (A - e \cdot b^T)]}{det(I - z \cdot A)} \cdot y_n$$

■

Nota 2.6 Para RK generales $det(I - z \cdot A), det[I - z \cdot (A - e \cdot b^T)]$ son polinomios de grado s, por lo que $R \in \prod_{s/s}$, aunque los grados del numerador o denominador pueden ser estrictamente menores que s. Por ejemplo, si $detA = 0$:

$$det(I - z \cdot A) = 1 + d_1 z + \cdots + det(A) \cdot (-1)^s \cdot z^s,$$

y el grado del denominador es menor o igual que $s - 1$.

Además para un RK explícito (A estrictamente triangular inferior) $det(I - z \cdot A) = det(I) = 1$, por lo que $R \in \prod_s$ (polinomio).

Las siguientes definiciones se introducen para caracterizar el conjunto de puntos z para los que la ecuación de prueba con paso h tiene un buen comportamiento de estabilidad.

Definición 2.8 Sea un método RK(A, b) con función de estabilidad lineal $R(z)$.

i) Se llama dominio de estabilidad absoluta S del método considerado a $S = \{z \in \mathbb{C}_\infty / |R(z)| \leq 1\}$
ii) El método se dice A-estable si $\mathbb{C}_\infty^- \subset S$, donde $\mathbb{C}_\infty^- = \{z \in \mathbb{C}_\infty / Rez \leq 0\}$
iii) El método se dice L-estable si es A-estable y $R(\infty) = 0$
iv) Sea $\alpha \in [0, \frac{\pi}{2}]$; el método se dice A($\alpha$)-estable si y solo si $\{z \in \mathbb{C}_\infty / |\pi - Argz| \leq \alpha\} \subset S$ (A-estable $\Leftrightarrow$ A($\frac{\pi}{2}$)-estable)
v) Un método se dice A_0-estable si $\{x \in \mathbb{R} : x \leq 0\} \subset S$
vi) Se llama intervalo de estabilidad absoluta $I \in A$ a la intersección de S con el eje real.

Hay métodos que no contienen a $\mathbb{C}_\infty^-$, pero si a buena parte de él. Por eso conviene introducir la A(α) estabilidad. Observemos que si $\hat{y}_{n+1} = y(t_{n+1}; 0, y_0)$ y $\hat{z}_{n+1} = y(t_{n+1}; 0, z_0)$ entonces $\hat{y}_{n+1} - \hat{z}_{n+1} = e^{\lambda \cdot t_{n+1}}(y_0 - z_0) = e^{(n+1)\cdot z}(y_0 - z_0)$, siendo $y_{n+1} - z_{n+1} = R(z) \cdot (y_n - z_n) = R(z)^{n+1} \cdot (y_0 - z_0)$. Y si $Re\lambda \to -\infty$ entonces $Rez \to -\infty$ y $e^z \to 0$. Los métodos A-estables para los que $R(z)$ reproduzca este comportamiento se dirán L-estables.

Nota 2.7 Tenemos que $\hat{y}_{n+1} = (e^z)^{n+1} \cdot y_0$ y que $y_{n+1} = (R(z))^{n+1} \cdot y_0$. Luego es de esperar cierta aproximación entre $R(z)$ y e^z.

Teorema 2.8. *Consideremos un RK(A, b) de orden p. Entonces* $e^z - R(z) = \mathcal{O}(z^{p+1}),\ z \to 0$.

Demostración. Puesto que se trata de un método de orden p, se tiene para el error local que

$$l(0, h) = \hat{y}_1 - \hat{y}_1 = (e^z - R(z)) \cdot y_0 = \mathcal{O}(z^{p+1}),\ \text{si } z \to 0 \quad (z = \lambda \cdot h).$$

Por tanto:

$$e^z - R(z) = \mathcal{O}(z^{p+1}),\ z \to 0.$$

■

Lema 2.2 Un RK(A, b) explícito y consistente no puede ser A-estable.

Demostración. Por consistencia: $b^T \cdot e = 1$. Así $p \geq 1$ y $R(z) - e^z = \mathcal{O}(z^2),\ z \to 0$. Como el método es explícito: $R(z) \in \prod_s$. Así $R(z) = 1 + z + a_2 z^2 + \cdots + a_s z^s$. En consecuencia, $|R(z)| \not\leq 1, \forall z \in \mathbb{C}^-$, y el método no es A-estable.

■

Como ya sabemos para un método RK de s etapas $R \in \prod_{s/s}$ y es A-estable si $|R(z)| \leq 1, \forall z \in \mathbb{C}^-$.

Definición 2.10 Una función racional R se dice A-aceptable si $|R(z)| \leq 1, \forall z \in \mathbb{C}^-$.

Lema 2.3 Un RK(A, b) es A-estable si y solo si $R(z)$ es A-aceptable.

Teorema 2.9. *Sea* $R \in \prod_{s/t}$. *R es A-aceptable si y solo si* $\begin{cases} |R(iy)| \leq 1, \forall y \in \mathbb{R} \\ R \text{ no tiene polos en } \mathbb{C}^- \end{cases}$

Demostración.
" $\Rightarrow$ " Trivial.
" $\Leftarrow$ " Supongamos que existe $z_0 \in \mathbb{C}^- : |R(z_0)| = \alpha > 1$.
Dado $r > |z_0|$, sea $S_r = \bar{D}(0, r) \bigcap \mathbb{C}^-$ y $\Gamma_r = \partial S_r$.
Como R no tiene polos en S_r (no los tiene en $\mathbb{C}^-$), por el principio del máximo: $\exists z_r \in \Gamma_r$ tal que $|R(z_r)| \geq \alpha > 1$. En particular, por hipótesis, $z_r \notin i \cdot \mathbb{R}$.
Haciendo $r \to \infty : |R(\infty)| = \lim\limits_{\substack{z \to \infty \\ z \in \mathbb{C}^-}} |R(z)| \geq \alpha > 1$. Pero, por otro lado, $\lim\limits_{\substack{z \to \infty \\ z = iy}} |R(z)| \leq 1$. Absurdo. Luego, $|R(z)| \leq 1, \forall z \in \mathbb{C}^-$.

■

Ejemplo 2.2 Consideremos el método de Gauss de una etapa dado por $\begin{array}{c|c} \frac{1}{2} & \frac{1}{2} \\ \hline & 1 \end{array}$
Entonces: $R(z) = \dfrac{1 + \frac{z}{2}}{1 - \frac{z}{2}}$.
R no tiene polos en $\mathbb{C}^-$ y $|R(iy)| = \dfrac{|1 + i\frac{y}{2}|}{|1 - i\frac{y}{2}|} = 1, \forall y \in \mathbb{R}$.
Luego, R es A-aceptable y el método A-estable. Además $R(\infty) = -1$.

Nota 2.8 La condición $|R(iy)| \leq 1, \forall y \in \mathbb{R}$, significa estabilidad sobre el eje imaginario, y se suele conocer como I-estabilidad. Esta condición es equivalente a que el polinomio $E(y) = |Q(iy)|^2 - |P(iy)|^2 = Q(iy) \cdot Q(-iy) - P(iy) \cdot P(-iy)$ satisfaga $E(y) \geq 0, \forall y \in \mathbb{R}$.

$$\begin{aligned} |R(iy)| \leq 1 &\Leftrightarrow |P(iy)| \leq |Q(iy)| \Leftrightarrow |P(iy)|^2 \leq |Q(iy)|^2 \\ &\Leftrightarrow P(iy) \cdot \overline{P(iy)} \leq Q(iy) \cdot \overline{Q(iy)} \\ &\Leftrightarrow P(iy) \cdot P(-iy) \leq Q(iy) \cdot Q(-iy) \\ &\Leftrightarrow E(y) \geq 0. \end{aligned}$$

Proposición 2.1

(a) $E(y)$ es un polinomio de grado par y $deg(E) \leq 2 \cdot \text{máx}(deg(P), deg(Q))$.
(b) Si $R(z)$ es una aproximación de orden p a e^z, entonces $E(y) = \mathcal{O}(y^{p+1}), y \to 0$.

Demostración.

(a) Puesto que $E(y) = E(-y)$ el grado de E debe ser par (es más, no contiene potencias impares de y). Que $deg(E) \leq 2 \cdot \text{máx}(deg(P), deg(Q))$ es inmediato.
(b) Puesto que R aproxima a e^z hasta el orden p:

$$e^z - R(z) = \mathcal{O}(z^{p+1}), z \to 0$$

Luego, $\frac{P(iy)}{Q(iy)} = e^{iy} + \mathcal{O}(y^{p+1}), y \to 0$.
Así:

$$\frac{|P(iy)|^2}{|Q(iy)|^2} = (e^{iy} + \mathcal{O}(y^{p+1})) \cdot (e^{-iy} + \mathcal{O}(y^{p+1})) = 1 + \mathcal{O}(y^{p+1})$$

De aquí, $E(y) = |Q(iy)|^2 - |P(iy)|^2 = \mathcal{O}(y^{p+1}), y \to 0$.

■

Proposición 2.2 Una función racional $R(z) = \frac{P(z)}{Q(z)} \in \prod_{s/t}$ de orden $p \geq 2t - 2$ es I-estable si y solo si $|R(\infty)| \leq 1$.

Demostración. " $\Rightarrow$ " Es evidente.
" $\Leftarrow$ " Ponemos $P(z) = p_0 + p_1 z + \cdots + p_s z^s$ y $Q(z) = q_0 + q_1 z + \cdots + q_t z^t$. Si $R(\infty) \leq 1$ entonces $s < t$ o si $s = t$, entonces $|p_s| \leq |q_t|$.
Por la *Proposición 2.1*: $E(y) = \mathcal{O}(y^{p+1}) = \mathcal{O}(y^{2n-1})$ y $E(y)$ no contiene potencias de y impares. Luego, como $s \leq t$ y $deg(E) \leq \text{máx}(s, t) = 2t$, se tiene que $E(y) = k \cdot y^{2t}, k \in \mathbb{R}$.
Ahora:

$$|P(iy)|^2 = P(iy) \cdot P(-iy) = 1 + \cdots + p_s^2 \cdot y^{2s}$$
$$|Q(iy)|^2 = Q(iy) \cdot Q(-iy) = 1 + \cdots + q_t^2 \cdot y^{2t}$$

por lo que:

$$E(y) = k \cdot y^{2t} = |Q(iy)|^2 - |P(iy)|^2 = (1-1) + \cdots - p_s^2 \cdot y^{2s} + q_t^2 \cdot y^{2t} = (q_t^2 - p_s^2 \cdot \delta_{st}) \cdot y^{2t},$$

siendo δ_{st} la delta de Kronecker. Así $k = q_t^2 - p_s^2 \cdot \delta_{st} \geq 0$.
En definitiva, $E(y) = k \cdot y^{2t} \geq 0, \forall y \in \mathbb{R}$ y $|R(iy)| \leq 1, \forall y \in \mathbb{R}$.

■

Corolario 2.2 Si $R(z) \in \prod_{s/t}$ es tal que no tiene polos en $\mathbb{C}^-$ y es de orden $p \geq 2t - 2$ entonces R es A-aceptable si y solo si $|R(\infty)| \leq 1$.

Finalmente presentamos un ejemplo en el que se estudia la estabilidad de los métodos SDIRK de 2 etapas. Estos métodos serán considerados posteriormente en el capítulo 3.

Ejemplo 2.3 Métodos SDIRK de dos etapas y orden $p \geq 2$.
Sean los métodos RK de la forma

$$\begin{array}{c|cc} c_1 & \gamma & 0 \\ c_2 & a_{21} & \gamma \\ \hline & b_1 & b_2 \end{array} \quad (\gamma > 0)$$

con $c_1 = \gamma$ y $c_2 = a_{21} + \gamma$. Imponiendo las condiciones de orden 2 (ver *Nota* 2.4)

$$\left.\begin{array}{l} b^T e = 1 \Leftrightarrow b_1 + b_2 = 1 \\ b^T c = \frac{1}{2} \Leftrightarrow b_1 c_1 + b_2 c_2 = \frac{1}{2} \end{array}\right\} \Leftrightarrow \left\{\begin{array}{l} b_1 = 1 - b_2 \\ b_2 = \frac{1-2\gamma}{2a_{21}} \end{array}\right. \quad (a_{21} \neq 0)$$

obtenemos una familia biparamétrica de métodos de dos etapas y orden $p \geq 2$, con parámetros $\gamma > 0$ y $a_{21} \neq 0$. Su función de estabilidad lineal $R(z) = \frac{det(I - z(A - eb^T))}{det(I - zA)}$ es de la forma

$$R(z) = \frac{P_2(z)}{(1-\gamma z)^2}, \text{ con } P_2(z) = 1 + \alpha_1 z + \alpha_2 z^2, \quad \alpha_1, \alpha_2 \in \mathbb{R},$$

y por tanto no tiene polos en $\mathbb{C}^-$. Luego, por el *Colorario 2.2*, el método es A-estable si y solo si $|R(\infty)| \leq 1$. Ahora bien,

$$R(\infty) = \frac{det(A - eb^T)}{det(A)} = 1 + \frac{1 - 4\gamma}{2\gamma^2}$$

y es sencillo comprobar que $-1 \leq R(\infty) \leq 1$ si y solo si $\gamma \geq \frac{1}{4}$. Por lo tanto, los métodos obtenidos son A-estables si y solo si $\gamma \geq \frac{1}{4}$. Por otra parte, $R(\infty) = 0 \Leftrightarrow \gamma = 1 \pm \frac{\sqrt{2}}{2}$. En ambos casos, $\gamma = 1 \pm \frac{\sqrt{2}}{2} > \frac{1}{4}$ y los métodos correspondientes son L-estables. Respecto a los coeficientes de error de orden 3 (ver *Nota* 2.4), resulta

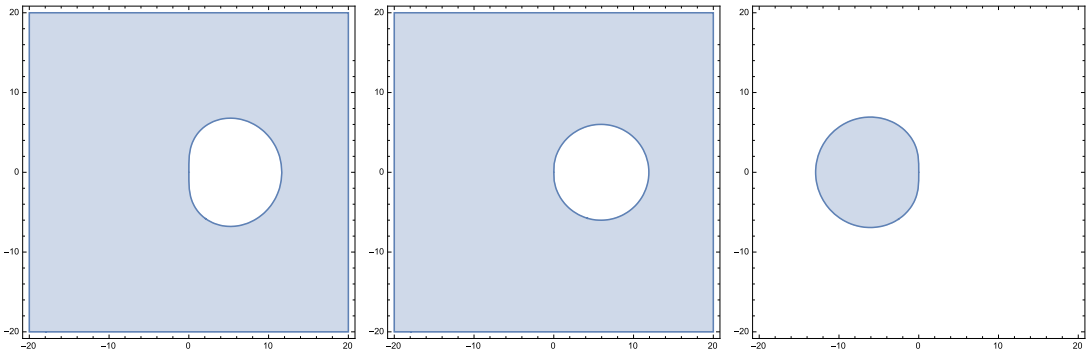
$$\begin{aligned} \tfrac{1}{3} - b^T c^2 &= \tfrac{1}{3} - a_{21}(\tfrac{1}{2} - \gamma) - \gamma + \gamma^2 \\ \tfrac{1}{6} - b^T A c &= \tfrac{1}{6} - \gamma + \gamma^2 \end{aligned}$$

y podemos tomar a_{21} a efectos de minimizar estos coeficientes de error como

$$\frac{1}{3} - a_{21}(\frac{1}{2} - \gamma) - \gamma + \gamma^2 = 0 \Leftrightarrow a_{21} = \frac{\frac{1}{3} - \gamma + \gamma^2}{\frac{1}{2} - \gamma} \text{ para } \gamma \neq \frac{1}{2}.$$

Observar que para $\gamma = \frac{1}{2}$, ambos coeficientes de error de orden 3 son independientes de a_{21}. Cuando $\gamma = \frac{3 \pm \sqrt{3}}{6}$ ambos coeficientes de error se anulan y se obtienen así dos métodos SDIRK de dos etapas y orden máximo $p = 3$ (ver *Teorema* 2.7). Si $\gamma = \frac{3 - \sqrt{3}}{6} \doteq 0.2113$ el método asociado no es A-estable, mientras que para $\gamma = \frac{3 + \sqrt{3}}{6} \doteq 0.7887$ el método correspondiente es A-estable con

$R(\infty) = 1 - \sqrt{3} \doteq -0.7321$.
En las gráficas siguientes, elaboradas usando el software Mathematica [11], se muestran los dominios de estabilidad lineal de los métodos SDIRK arriba descritos para $\gamma = 1 - \frac{\sqrt{2}}{2}$ (orden 2, L-estable), $\gamma = \frac{3+\sqrt{3}}{6}$ (orden 3, A-estable con $-1 < R(\infty) < 0$) y $\gamma = \frac{3-\sqrt{3}}{6}$ (orden 3, no A-estable). La región sombreada representa el conjunto de puntos z donde $|R(z)| \leq 1$. Estos tres métodos serán considerados en el próximo capítulo de cara a la integración temporal de las ecuaciones de Navier-Stokes incompresibles.



(a) $\gamma = 1 - \frac{\sqrt{2}}{2}$ (b) $\gamma = \frac{3+\sqrt{3}}{6}$ (c) $\gamma = \frac{3-\sqrt{3}}{6}$

Figura 2.1: Regiones de estabilidad para métodos SDIRK de dos etapas.

3

Métodos de proyección para ecuaciones de Navier-Stokes incompresibles

En este último capítulo vamos a tratar una alternativa de resolución numérica de las ecuaciones de Navier-Stokes para fluidos incompresibles. Procederemos a realizar una discretización temporal en las ecuaciones mediante un método de tipo Runge-Kutta (nosotros trabajaremos con Euler implícito y SDIRK de dos etapas), luego en un paso intermedio calcularemos una velocidad estimada con la cual podremos obtener finalmente la solución para la presión y la velocidad resolviendo ciertas ecuaciones en derivadas parciales de tipo elíptico (mediante métodos de elementos finitos). Esta idea es lo que se conoce como el método de proyección de Chorin [4]. Por último ilustraremos en varios ejemplos la implementación del método.

Recordemos las ecuaciones de Navier-Stokes para fluidos incompresibles

$$\begin{cases} \mathbf{u}_t - \nu\Delta\mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u} + \nabla p = f \\ \text{div } \mathbf{u} = 0 \end{cases} \qquad (\mathbf{x},t) \in \Omega \times (0,T], \tag{3.1}$$

con la condición de frontera

$$\mathbf{u}(\mathbf{x},t) = 0, \qquad (\mathbf{x},t) \in \partial\Omega \times (0,T],$$

y la condición inicial

$$\mathbf{u}(\mathbf{x},0) = \mathbf{u}_0(\mathbf{x}), \qquad \mathbf{x} \in \Omega.$$

Definiendo

$$w := f + \nu\Delta\mathbf{u} - (\mathbf{u}\cdot\nabla)\mathbf{u},$$

por (3.1) tenemos que

$$w = \mathbf{u}_t + \nabla p \quad \text{con} \quad \begin{cases} \text{div } (\mathbf{u}_t) = (\text{div } \mathbf{u})_t = 0, & \text{en } \Omega \times (0,T] \\ \mathbf{u}_t \cdot \mathbf{n} = (\mathbf{u}\cdot\mathbf{n})_t = 0, & \text{en } \partial\Omega \times (0,T] \end{cases}$$

Luego, $\forall t > 0, w = \mathbf{u}_t + \nabla p$ corresponde a una descomposición de Helmholtz-Hodge (1.11) para w y por tanto:

$$\begin{cases} \Delta p = \text{div } w \quad \text{en } \Omega \\ \dfrac{\partial p}{\partial n} = w \cdot n \quad \text{en } \partial\Omega \end{cases} , \forall t \in (0, T].$$

Esto define p, salvo constante aditiva, a partir de $\mathbf{u}$. Por este motivo no son necesarias condiciones iniciales ni de frontera para p en las ecuaciones (3.1).

3.1. Método de Chorin con Euler Implícito

Expresando la ecuación (3.1) como

$$\mathbf{u}_t = g(\mathbf{u}) := \nu\Delta\mathbf{u} - (\mathbf{u} \cdot \nabla)\mathbf{u} - \nabla p + f$$

podemos aplicar el método de Euler implícito como un método de semi-discretización en el tiempo

$$\mathbf{u}^{n+1} = \mathbf{u}^n + \Delta t \cdot g(\mathbf{u}^{n+1}) = \mathbf{u}^n + \Delta t \cdot [\nu\Delta\mathbf{u}^{n+1} - (\mathbf{u}^{n+1} \cdot \nabla)\mathbf{u}^{n+1} - \nabla p^{n+1} + f^{n+1}]$$
$$\text{div } \mathbf{u}^{n+1} = 0$$

donde $g(\mathbf{u}^{n+1}) = K_1$. Esto nos llevará a resolver las siguientes ecuaciones

$$\begin{cases} \dfrac{\mathbf{u}^{n+1} - \mathbf{u}^n}{\Delta t} = \nu\Delta\mathbf{u}^{n+1} - (\mathbf{u}^{n+1} \cdot \nabla)\mathbf{u}^{n+1} - \nabla p^{n+1} + f^{n+1} \\ \text{div } (\mathbf{u}^{n+1}) = 0 \end{cases} , n \geq 0. \quad (3.2)$$

en cada tiempo de la partición $0 = t_0 < \ldots < t_{n+1} < \ldots < t_{N+1} \equiv T$.

Si aplicasemos una discretización espacial en todo tiempo t_{n+1}, aparecerían sistemas de ecuaciones lineales de gran complejidad. Además, es necesario una reducción significativa del esfuerzo computacional del algoritmo para el cómputo de fluidos complejos. Por ello se han propuesto métodos numéricos cuyo objetivo es calcular la tupla $\{\mathbf{u}^{n+1}, p^{n+1}\}$ en *pasos separados* lo que conlleva a una drástica reducción del costo computacional. Estos son conocidos como métodos de proyección y el primero de ellos fue formulado por Chorin en 1968 ([4],[10]) tal como sigue:

1. Empezar suponiendo que $\mathbf{u}^0 \approx \mathbf{u}_0$.
2. Para $n \geq 0$, suprimir el término de la presión ∇p y definir

$$\hat{g}(\mathbf{u}) = \nu\Delta\mathbf{u} - (\mathbf{u} \cdot \nabla)\mathbf{u} + f.$$

Hallar $\tilde{\mathbf{u}}^{n+1}$ resolviendo

$$\tilde{\mathbf{u}}^{n+1} = \mathbf{u}^n + \Delta t \cdot \hat{K}_1 = \mathbf{u}^n + \Delta t \cdot \hat{g}(\tilde{\mathbf{u}}^{n+1}), \quad \tilde{\mathbf{u}}^{n+1}|_{\partial\Omega} = 0. \quad (3.3)$$

Esto equivale a resolver

$$\hat{K}_1 = \hat{g}(\mathbf{u}^n + \Delta t \cdot \hat{K}_1), \quad \text{con} \quad \hat{K}_1|_{\partial\Omega} = 0$$

y hallado $\hat{K}_1$, se computa

$$\tilde{\mathbf{u}}^{n+1} = \mathbf{u}^n + \Delta t \cdot \hat{K}_1.$$

3. Una vez tenemos $\tilde{\mathbf{u}}^{n+1}$, determinaremos la solución de la tupla $\{\mathbf{u}^{n+1}, p^{n+1}\}$ resolviendo

$$\begin{cases} \dfrac{\mathbf{u}^{n+1} - \tilde{\mathbf{u}}^{n+1}}{\Delta t} + \nabla p^{n+1} = 0 \\ \text{div}\,(\mathbf{u}^{n+1}) = 0 \end{cases} , \mathbf{u}^{n+1}|_{\partial\Omega} \cdot \mathbf{n} = 0. \tag{3.4}$$

Para ello se computa p^{n+1} aplicando divergencia en (3.4)

$$\begin{cases} -\Delta p^{n+1} = -\dfrac{1}{\Delta t} \cdot \text{div}(\tilde{\mathbf{u}}^{n+1}) \\ \left(\dfrac{\mathbf{u}^{n+1} - \tilde{\mathbf{u}}^{n+1}}{\Delta t}\right)\Big|_{\partial\Omega} \cdot \mathbf{n} = -\nabla p^{n+1}|_{\partial\Omega} \cdot \mathbf{n} = \partial_n p^{n+1}|_{\partial\Omega} = 0, \end{cases} \tag{3.5}$$

que corresponde a un problema de Poisson-Neumann para p^{n+1}. Hallado p^{n+1}, podemos calcular el campo de velocidades con la siguiente actualización

$$\mathbf{u}^{n+1} = \tilde{\mathbf{u}}^{n+1} - \Delta t \cdot \nabla p^{n+1}.$$

Luego, cada iteración se divide en dos problemas elementales: una ecuación de convección-difusión con condiciones de Dirichlet para el campo $\tilde{\mathbf{u}}^{n+1}$ (3.3) y una ecuación de Poisson con condiciones de Neumann, para el computo de la presión p^{n+1} (3.5).

A continuación introduciremos dos resultados sobre el orden de convergencia y los errores globales en norma euclídea para las soluciones del método de Chorin basado en el método de Euler (véase [10, Cap. 6] para más detalles).

Teorema 3.1. *Sea $\{\boldsymbol{u}^{n+1}, p^{n+1}\}$ la solución (semi-)discreta del método de Chorin (3.3)-(3.5), y $\{\boldsymbol{u}(t_{n+1}), p(t_{n+1})\}$ la solución de las ecuaciones de Navier-Stokes (3.1) en tiempo $0 < t_{n+1} \leq T \equiv t_{N+1}$. Asumiendo suficiente regularidad en el dominio así como en los datos del problema y asumiendo que el problema de Navier-Stokes está bien planteado, es decir, existe una única solución $\{\boldsymbol{u}, p\}$ definida globalmente en $[0, T]$, para tamaños de paso suficientemente pequeños $\Delta t \leq k_0(T)$ existe una constante C, que solo depende de los datos del problema, de modo que se cumplen las siguientes estimaciones*

i) $\max\limits_{0\le n\le N} \|\boldsymbol{u}(t_{n+1}) - \boldsymbol{u}^{n+1}\| \le C\Delta t.$

ii) $\max\limits_{0\le n\le N} \sqrt{\tau_{n+1}}\,\|p(t_{n+1}) - p^{n+1}\| \le C\sqrt{\Delta t},$ *siendo* $\tau_{n+1} = \min\{1, t_{n+1}\}$

Teorema 3.2. *Bajo las mismas condiciones que en el Teorema 3.1, se cumplen las siguientes estimaciones del error local para subdominios interiores* $\Omega' \subset \Omega$

i) $\max\limits_{0\le n\le N} \|\boldsymbol{u}(t_{n+1}) - \boldsymbol{u}^{n+1}\|_{\Omega'} \le \tilde{C}\Delta t.$

ii) $\max\limits_{0\le n\le N} \sqrt{\tau_{n+1}}\,\|p(t_{n+1}) - p^{n+1}\|_{\Omega'} \le \tilde{C}\Delta t.$

Nota 3.1 La norma euclídea para $\mathbf{v}$ una función de cuadrado integrable en Ω' es $\|\mathbf{v}\| = (\int_{\Omega'} \|\mathbf{v}\|_2^2 dx)^{\frac{1}{2}}$. Los dos teoremas previos aseguran orden de convergencia 1 en norma euclídea para el método (3.3)-(3.5) en la componente $\mathbf{u}$, y al menos orden $\frac{1}{2}$ en la presión p (con orden 1 en subdominios).

3.2. Método de Chorin con métodos SDIRK

Para aplicar el método SDIRK de dos etapas procederemos de manera análoga a como hemos hecho con Euler implícito.

1. Suponemos que $\mathbf{u}^0 \approx \mathbf{u}_0$
2. Para $n \ge 0$, hallamos $\tilde{\mathbf{u}}^{n+1}$ como sigue. Definimos etapas $\hat{K}_1$ y $\hat{K}_2$ como

$$\begin{cases} \hat{K}_1 = \hat{g}(\mathbf{u}^n + \gamma\Delta t \cdot \hat{K}_1), \qquad \hat{K}_1|_{\partial\Omega} = 0. \\ \hat{K}_1 = \nu\Delta y - (y\cdot\nabla)y + f^{n+1}, \quad \text{con} \quad y = \mathbf{u}^n + \gamma\Delta t \cdot \hat{K}_1. \end{cases}$$

$$\begin{cases} \hat{K}_2 = \hat{g}(\mathbf{u}^n + a_{21}\Delta t \cdot \hat{K}_1 + \gamma\Delta t \cdot \hat{K}_2), \qquad \hat{K}_2|_{\partial\Omega} = 0. \\ \hat{K}_2 = \nu\Delta y - (y\cdot\nabla)y + f^{n+1}, \quad \text{con} \quad y = \mathbf{u}^n + a_{21}\Delta t \cdot \hat{K}_1 + \gamma\Delta t \cdot \hat{K}_2. \end{cases}$$

Una vez halladas $\hat{K}_1$ y $\hat{K}_2$ la solución (proyectada) del SDIRK será:

$$\tilde{\mathbf{u}}^{n+1} = \mathbf{u}^n + \Delta t[b_1 \cdot \hat{K}_1 + b_2 \cdot \hat{K}_2].$$

3. Con la solución proyectada $\tilde{\mathbf{u}}^{n+1}$ para obtener $\{\mathbf{u}^{n+1}, p\}$ se procede de manera análoga a como hemos hecho para Euler implícito considerando (3.4)-(3.5).

Nota 3.2 Recordemos que los coeficientes del método SDIRK de dos etapas y orden $p \ge 2$ (con coeficientes de error de orden 3 mínimos para $p = 2$) vienen dados por

$$a_{21} = \frac{2 + 6\gamma(\gamma - 1)}{3(1 - 2\gamma)}, \; b_1 = \frac{1}{4(1 - 3\gamma + 3\gamma^2)}, \; b_2 = 1 - b_1 \quad (\gamma \ne \frac{1}{2}).$$

- Orden 3: $\gamma = \dfrac{3 \pm \sqrt{3}}{6}$

  $\gamma = \dfrac{3 - \sqrt{3}}{6} = 0.21\ldots$ no A-estable, $\gamma = \dfrac{3 + \sqrt{3}}{6}$ A-estable $(R(\infty) \doteq -0.73)$

- Orden 2 y L-estabilidad: $\gamma = \dfrac{2 \pm \sqrt{2}}{2}$

  Tomaremos $\gamma = \dfrac{2 - \sqrt{2}}{2} \simeq 0.29$ para obtener coeficientes de error con menor valor absoluto.

3.3. Ilustración numérica

Para concluir este trabajo se realizarán una serie de experimentos numéricos en los que se resolverán las ecuaciones de Navier-Stokes para fluidos incompresibles aplicando el método de Chorin. Esto se hará mediante una implementación en *Python* y con la ayuda del software FEniCS ([1],[9]) que es capaz de resolver ecuaciones en derivadas parciales mediante métodos de elementos finitos ([8]). En todos los casos que estudiaremos se considera velocidad inicial $\mathbf{u}_0 = 0$ y se establecen condiciones *no-slip* en la frontera, es decir, $\mathbf{u}(\mathbf{x}, t)|_{\partial\Omega} = 0, \forall t$. Se mostrarán las gráficas obtenidas en tiempo final de la solución de $\mathbf{u}$ y p para cada problema considerado. Además señalar que para el método SDIRK se usarán los coeficientes de la *Nota 3.2*.

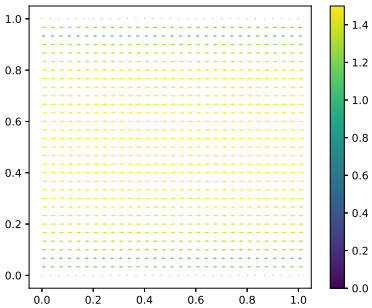
1. En el primer ejemplo el dominio Ω será el cuadrado unidad y el fluido es empujado de izquierda a derecha por una fuerza constante $f = (1, 0)^T$. El coeficiente de viscosidad establecido es $\nu = 0.01$, el tamaño de paso $\Delta t = 0.1$ y el tiempo final $T = 1.5$. Se añade la condición de $p|_{\partial\Omega} = 0$ necesaria para computar un valor concreto de la presión en la ecuación de Poisson-Neumann del segundo paso en el método de Chorin. Aplicaremos el método de Euler implícito y los métodos SDIRK de dos etapas. Observamos en la figuras 3.1g y 3.1h que para métodos no A-estables pueden aparecer perturbaciones en la solución.

2. En segundo lugar aplicaremos los métodos anteriormente citados en un dominio Ω que simule una tubería en forma de L. Para ello tomaremos el subconjunto que se obtiene al suprimir el cuadrante superior derecho del cuadrado unidad. Además en este caso tomaremos un coeficiente de viscosidad más pequeño $\nu = 0.001$ lo que conlleva a que el problema sea menos difusivo, es decir, convección (transporte) más dominante. Tomaremos un paso de tiempo $\Delta t = 0.05$, tiempo final $T = 1.5$ y no habrá fuerzas externas $f = (0, 0)^T$, sino que aplicaremos las condiciones de presión $p_{in} = 1$ para el fluido que

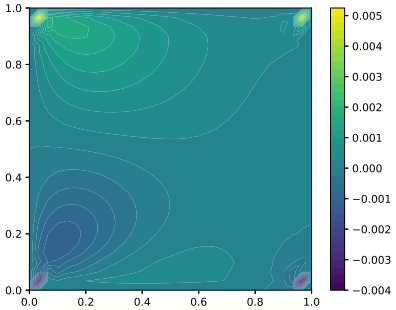
entra cuando $y = 1$ y $p_{out} = 0$ para el fluido que sale cuando $x = 1$. Esto hará que el fluido baje y siga por la tubería hacia la derecha.

3. El último ejemplo consiste en aplicar los métodos en un dominio $\Omega = [0,1]^2 \setminus D$, donde D adopta la forma de un delfín. Se considera una viscosidad $\nu = 0.01$, un tamaño de paso $\Delta t = 0.05$ y un tiempo final $T = 0.3$. La fuerza externa $f = (1,0)^T$ empujará al fluido de izquierda a derecha. Se aplica la condición para la presión en la frontera $p|_{\partial\Omega} = 0$. Se vuelven a observar perturbaciones en la solución cuando el método no es A-estable.

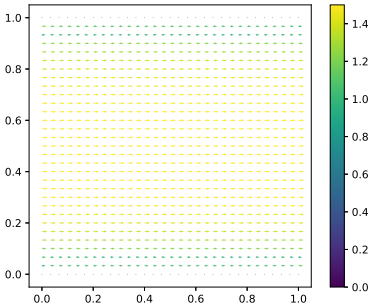
Nota 3.3 Se ilustrará la implementación de estos tres ejemplos mediante las gráficas de la solución para cada método numérico tanto de la velocidad (columna izquierda) como de la presión (columna derecha) en sus respectivos tiempo finales.



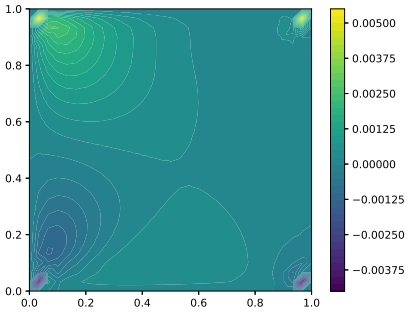
(a) Euler implícito



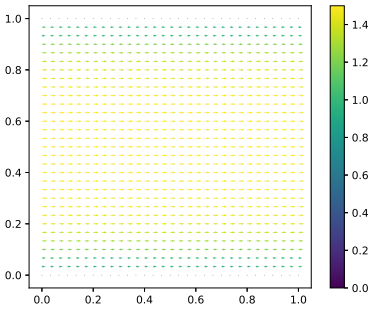
(b) Euler implícito



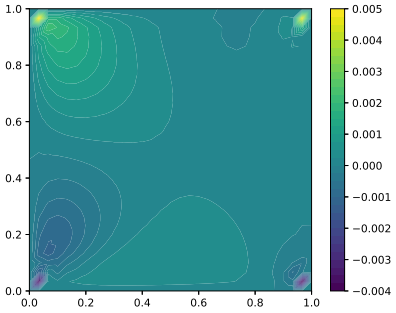
(c) SDIRK, Orden 2, L-estable



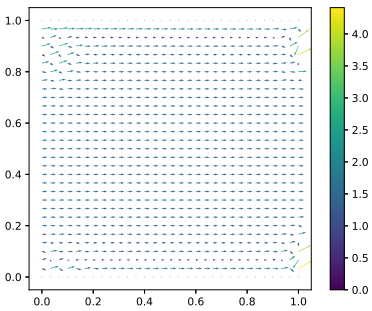
(d) SDIRK, Orden 2, L-estable



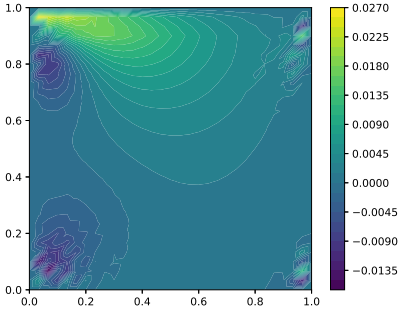
(e) SDIRK, Orden 3, A-estable



(f) SDIRK, Orden 3, A-estable

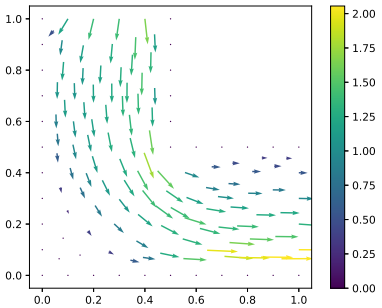


(g) SDIRK, Orden 3, no A-estable

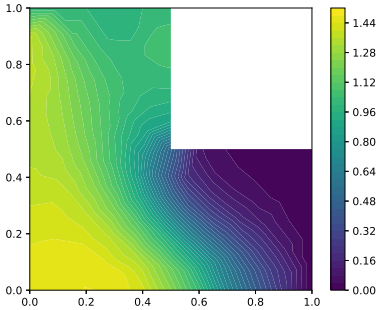


(h) SDIRK, Orden 3, no A-estable

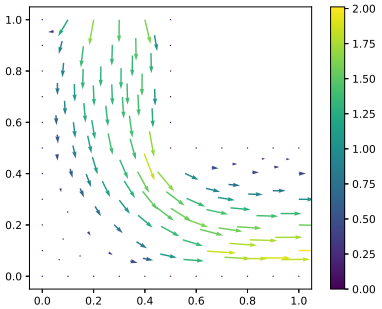
Figura 3.1: Soluciones de la velocidad $\mathbf{u}$ (izquierda) y la presión p (derecha) en el *Ejemplo 1.*



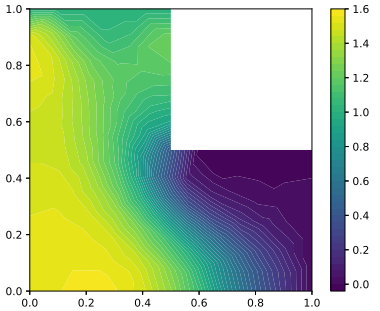
(a) Euler implícito



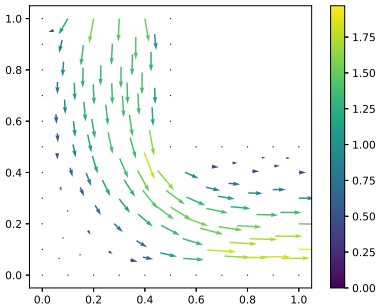
(b) Euler implícito



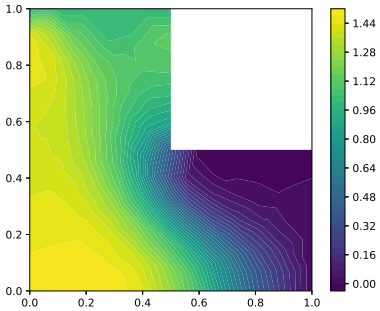
(c) SDIRK, Orden 2, L-estable



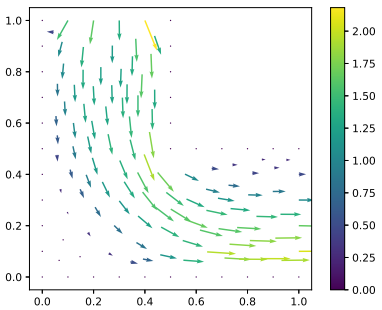
(d) SDIRK, Orden 2, L-estable



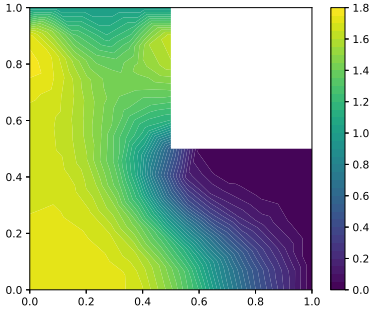
(e) SDIRK, Orden 3, A-estable



(f) SDIRK, Orden 3, A-estable

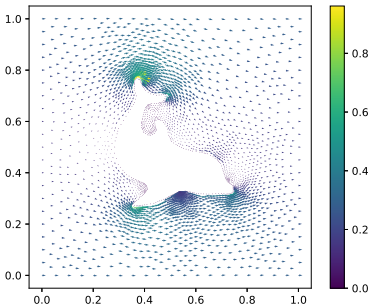


(g) SDIRK, Orden 3, no A-estable

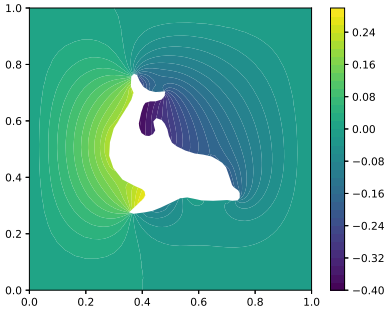


(h) SDIRK, Orden 3, no A-estable

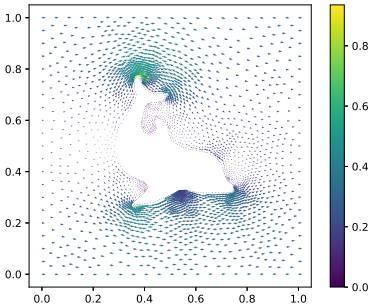
Figura 3.2: Soluciones de la velocidad $\mathbf{u}$ (izquierda) y la presión p (derecha) en el *Ejemplo 2*.



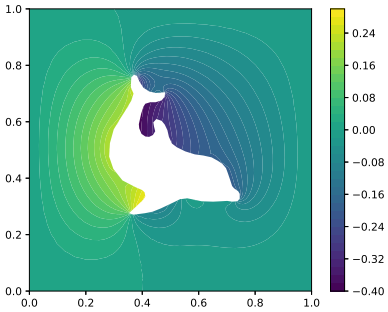
(a) Euler implícito



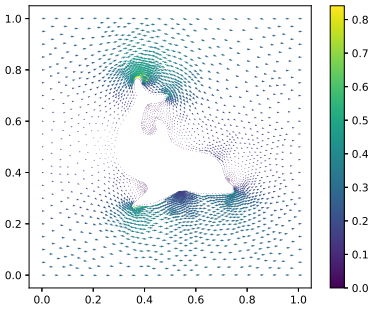
(b) Euler implícito



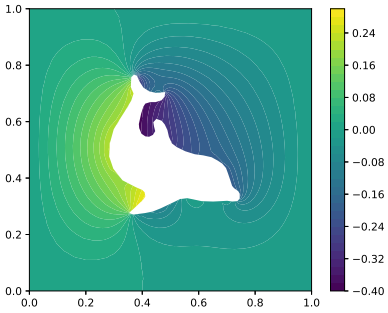
(c) SDIRK, Orden 2, L-estable



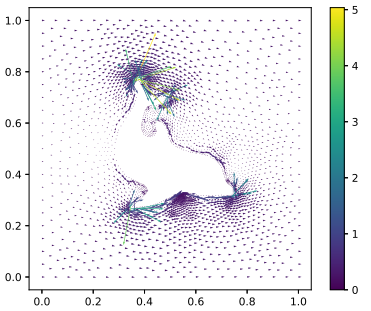
(d) SDIRK, Orden 2, L-estable



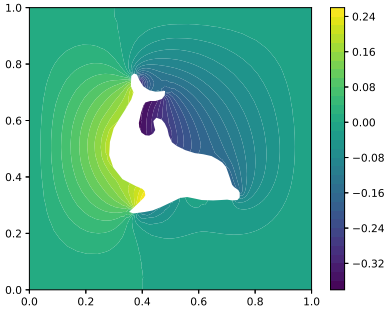
(e) SDIRK, Orden 3, A-estable



(f) SDIRK, Orden 3, A-estable



(g) SDIRK, Orden 3, no A-estable



(h) SDIRK, Orden 3, no A-estable

Figura 3.3: Soluciones de la velocidad $\mathbf{u}$ (izquierda) y la presión p (derecha) en el *Ejemplo 3*.

Bibliografía

[1] **M. S. Alnaes, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes and G. N. Wells**, *FEniCS: The FEniCS Project Version 1.5.* Archive of Numerical Software, vol. 3, 2015. https://fenicsproject.org/.

[2] **J.C. Butcher,** *Numerical methods for Ordinary Differential Equations*, John Wiley (2008).

[3] **M. Calvo, J.I. Montijano, L. Rández**, *Métodos de Runge–Kutta para la resolución numérica de ecuaciones diferenciales ordinarias*, Publicaciones de la Universidad de Zaragoza (1984).

[4] **A.J. Chorin**, *Numerical solution of the Navier-Stokes equations*, Math. Comp. 104 (vol. 22), p. 745-762 (1968).

[5] **A.J. Chorin, J.E. Marsden**, *A mathematical introduction to fluid mechanics*, Springer (2000).

[6] **E. Hairer, S.P. Nørsett, G. Wanner**, *Solving Ordinary Differential Equations I, Nonstiff problems*, Springer (2008).

[7] **E. Hairer, G. Wanner**, *Solving Ordinary Differential Equations II, Stiff problems*, Springer (2002).

[8] **S. Larsson, V. Thomée**, *Partial Differential Equations with Numerical Methods*, Springer, 2009.

[9] **A. Logg and G. N. Wells**, *Dolphin: Automated Finite Element Computing.* ACM Transactions on Mathematical Software, vol. 37, 2010.

[10] **A. Prohl**, *Projection and quasi-compressibility methods for solving the incompressible Navier-Stokes equations*, Springer (1997).

[11] **S. Wolfram**, *The Mathematica book* 2003 (5th. ed.), Wolfram Media Inc.

SDIRK methods for incompressible Navier-Stokes equations

Sección de Matemáticas
Universidad de La Laguna

Romen Santana Benítez

Facultad de Ciencias · Sección de Matemáticas
Universidad de La Laguna
alu0100990942@ull.edu.es

Abstract

IN THIS MANUSCRIPT *we are going to deduce the Navier-Stokes partial differential equations starting from three fundamental principles of fluid mechanics: conservation of mass, Newton's second law (balance of momentum) and conservation of energy. Then we will proceed to introduce the Runge-Kutta numerical methods for ordinary differential equations, studying their consistency, stability and convergence. This will be necessary because we will end up implementing what is known as projection methods, specifically the Chorin method, considering simply diagonally implicit Runge-Kutta methods (SDIRK). These numerical methods will allow us to approximate the solution of the Navier-Stokes equations for incompressible fluids. Finally, several examples of the Chorin's method implementation will be illustrated, which has been carried out with the help of Python and the FEniCS software.*

1. Incompressible Navier-Stokes equations

FROM the principles of fluids mechanics we can introduce the Euler's equations for an incompressible ideal flow

$$\begin{cases} \rho\dfrac{D\mathbf{u}}{Dt} = -\nabla p + \rho\mathbf{b} \\ \dfrac{D\rho}{Dt} = 0 \\ \text{div } \mathbf{u} = 0. \end{cases}$$

For more general fluids we obtain what is known as the Navier-Stokes equations for incompressible flow

$$\begin{cases} \mathbf{u}_t - \nu\Delta\mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u} + \nabla p = f \\ \text{div } \mathbf{u} = 0 \end{cases} \quad (\mathbf{x}, t) \in \Omega \times (0, T].$$

These equations are supplemented by boundary conditions. For Euler's equations for ideal flow we use $\mathbf{u}|_{\partial\Omega}\cdot\mathbf{n} = 0$. For the Navier-Stokes equations, the extra term $\nu\Delta\mathbf{u}$ raises the number of derivatives of $\mathbf{u}$ involved from one to two, this is accompanied by an increase in the number of boundary conditions.

2. Runge-Kutta and SDIRK methods

A Runge-Kutta method of s stages applied to an IVP with time step $h > 0$ from $t = t_0$ to $t_1 = t_0 + h$ computing $y_1 \simeq y(t_0 + h)$ is defined by:

$$\begin{cases} K_i = f(t_0 + c_i\cdot h, y_0 + h\cdot\sum_{j=1}^{s} a_{ij}K_j),\ 1 \le i \le s, \\ y_1 = y_0 + h\cdot\sum_{i=1}^{s} b_i K_i. \end{cases}$$

The two stage SDIRK methods with order $p \ge 2$ that are going to be used in this work are Runge-Kutta type methods with the shape

$$\begin{array}{c|cc} c_1 & \gamma & 0 \\ c_2 & a_{21} & \gamma \\ \hline & b_1 & b_2 \end{array} \quad (\gamma > 0),$$

where $c_1 = \gamma$ and $c_2 = a_{21} + \gamma$. We establish the order 2 and order 3 conditions for the method:

$$\begin{cases} b^T e = 1 \\ b^T c = \frac{1}{2} \end{cases} \text{ and } \begin{cases} b^T c^2 = \frac{1}{3} \\ b^T A c = \frac{1}{6} \end{cases}$$

in order to obtain the following coefficients: $\gamma = \frac{2-\sqrt{2}}{2}$ (order 2, L-stable), $\gamma = \frac{3+\sqrt{3}}{6}$ (order 3, A-stable) and $\gamma = \frac{3-\sqrt{3}}{6}$ (order 3, not A-stable).

3. Chorin's projection method

THE application of the implicit Euler method as a semi-discretization scheme in time leads to the following equations to be solved

$$\begin{cases} \dfrac{\mathbf{u}^{n+1} - \mathbf{u}^n}{\Delta t} = \nu\Delta\mathbf{u}^{n+1} - (\mathbf{u}^{n+1}\cdot\nabla)\mathbf{u}^{n+1} - \nabla p^{n+1} + f^{n+1} \\ \text{div } (\mathbf{u}^{n+1}) = 0 \end{cases}, n \ge 0.$$

Projection schemes compute the tuple $\{\mathbf{u}^{n+1}, p\}$ in separate steps, this yields a drastic reduction of computational work. The firs projection method was formulated by Chorin in 1968 and is on the following form:

1. Start with an appropriate initial guess $\mathbf{u}^0 \approx \mathbf{u}_0$.
2. For $n \ge 0$, find $\tilde{\mathbf{u}}^{n+1}$ as the solution of (we get rid of the pressure)
$$\begin{cases} \dfrac{\tilde{\mathbf{u}}^{n+1} - \mathbf{u}^n}{\Delta t} - \nu\Delta\tilde{\mathbf{u}}^{n+1} + (\tilde{\mathbf{u}}^{n+1}\cdot\nabla)\tilde{\mathbf{u}}^{n+1} = f^{n+1} \\ \tilde{\mathbf{u}}^{n+1}|_{\partial\Omega} = 0. \end{cases}$$
3. Provided with $\tilde{\mathbf{u}}^{n+1}$, determine the solution of the tuple $\{\mathbf{u}^{n+1}, p\}$ as the solution of
$$\begin{cases} \dfrac{\mathbf{u}^{n+1} - \tilde{\mathbf{u}}^{n+1}}{\Delta t} + \nabla p^{n+1} = 0 \\ \text{div } (\mathbf{u}^{n+1}) = 0. \end{cases}, \ \mathbf{u}^{n+1}|_{\partial\Omega}\cdot\mathbf{n} = 0.$$

4. Numerical ilustration

FINALLY we implement the Chorin's method with Python and *FEniCS* software. This is an example of the solutions obtained with the Chorin's scheme using a two-stage SDIRK discretization method, specifically an order 3 A-stable method with $\gamma = \frac{3+\sqrt{3}}{6}$. The domain Ω has an L-pipe shape.

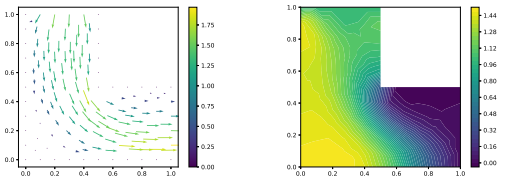


Figure 1: *Velocity* $\mathbf{u}$ *(left side) and pressure* p *(right side).*

References

[1] **A.J. Chorin, J.E. Marsden**, *A mathematical introduction to fluid mechanics*, Springer (2000).

[2] **M. Calvo, J.I. Montijano, L. Rández**, *Métodos de Runge–Kutta para la resolución numérica de ecuaciones diferenciales ordinarias*, Publicaciones de la Universidad de Zaragoza (1984).

[3] **A. Prohl**, *Projection and quasi-compressibility methods for solving the incompressible Navier-Stokes equations*, Springer (1997).

[4] **M. S. Alnaes, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes and G. N. Wells**, *FEniCS: The FEniCS Project Version 1.5.* Archive of Numerical Software, vol. 3, 2015. https://fenicsproject.org/.

TRABAJO FIN DE GRADO, Convocatoria de Junio, 2021