Finally we note that with the present diffusion-reaction problem the reaction term is not stiff, so this could also be treated explicitly. However, with the Douglas method (3.6) this would mean in a forward Euler fashion, which will lead therefore to first-order convergence only. Splitting type methods with improved treatment of explicit terms are studied in the next two sections.

4 IMEX Methods

For many problems there are natural splittings into two parts, one of which is non-stiff, or mildly stiff, and suited for explicit treatment. This can be handled in a straightforward way within operator splitting but then accuracy and boundary conditions are a major concern. Moreover, with operator splitting, multistep methods are not suited to solve the fractional steps, because if we solve a subproblem $v'(t) = F_i(t, v(t))$ on $[t_n, t_{n+1}]$, a multistep method will need past information for that particular subproblem, rather than values w_{n-j} for the whole problem.

In this section we consider IMEX methods, which consist of suitable mixtures of implicit and explicit methods. There exist IMEX methods of linear multistep and Runge-Kutta type. Here we will mainly focus on linear two-step methods. First we illustrate the ideas through the one-step IMEX-θ method.

4.1 The IMEX-θ Method

Suppose that the semi-discrete system is given by

$$w'(t) = F(t, w(t)) \equiv F_0(t, w(t)) + F_1(t, w(t)), \tag{4.1}$$

where F_0 is a non-stiff term suitable for explicit time integration, for instance discretized advection, and F_1 is a stiff term requiring an implicit treatment, say discretized diffusion or stiff reactions. Consider the following simple method

$$w_{n+1} = w_n + \tau F_0(t_n, w_n) + (1-\theta)\tau F_1(t_n, w_n) + \theta\tau F_1(t_{n+1}, w_{n+1}), \tag{4.2}$$

with parameter $\theta \geq \frac{1}{2}$. Here one sees that the explicit Euler method is combined with the A-stable implicit θ-method. Such mixtures of implicit and explicit methods are called IMEX methods. Method (4.2) is called the IMEX-θ method.

Inserting the exact solution of (4.1) gives the temporal truncation error

$$\begin{aligned}\rho_n &= \tau^{-1}\big(w(t_{n+1}) - w(t_n)\big) - (1-\theta)F(t_n, w(t_n)) - \theta F(t_{n+1}, w(t_{n+1})) \\ &\quad + \theta\big(\varphi(t_{n+1}) - \varphi(t_n)\big) = (\tfrac{1}{2} - \theta)\tau w''(t_n) + \theta\tau\varphi'(t_n) + \mathcal{O}(\tau^2),\end{aligned}$$

where $\varphi(t) = F_0(t, w(t))$. If F_0 represents discretized advection or other non-stiff terms, smoothness of w will usually also imply smoothness of φ, independent of boundary conditions or small mesh widths h. Therefore the structure of the truncation error is much more favourable than for methods based on operator splitting with fractional steps. For example, with a stationary solution we have a zero truncation error. On the other hand, with methods of this IMEX type it is stability that needs a careful examination.

The above method (4.2) is the most simple IMEX method. In fact this method can be viewed as a special case ($s = 1$) of the Douglas method (3.6), which has been studied already in Section 3.2 but mainly with $F_0 = 0$ and multiple implicit terms.

Stability

As before, let us consider the scalar complex test equation

$$w'(t) = \lambda_0 w(t) + \lambda_1 w(t)\,, \tag{4.3}$$

and let $z_j = \tau\lambda_j$, $j = 0, 1$. Recall that in applications to PDEs these λ_j represent, after linearization, eigenvalues of the two components F_0 and F_1 found by inserting Fourier modes. One would hope that having $|1 + z_0| \le 1$ (stability of the explicit method) and $\mathrm{Re}\,(z_1) \le 0$ (stability of the implicit method) would be sufficient for linear stability of the IMEX method, but in general this is not true. Application of (4.2) to this test equation yields $w_{n+1} = R\,w_n$, $R = R(z_0, z_1)$, with

$$R(z_0, z_1) = \frac{1 + z_0 + (1-\theta)z_1}{1 - \theta z_1} \tag{4.4}$$

and stability for the test equation thus requires $|R(z_0, z_1)\,| \le 1$.

We first consider the set

$$\mathcal{D}_0 = \{z_0 \in \mathbb{C} : \text{the IMEX scheme is stable for any } z_1 \in \mathbb{C}^-\}\,. \tag{4.5}$$

So here we insist on A-stability with respect to the implicit part, and the question is whether $\mathcal{D}_0$ is smaller than the stability region of the explicit Euler method. Considering $z_1 = it$, $t \in \mathbb{R}$, and using the maximum modulus principle, it follows by some straightforward calculations that $z_0 = x_0 + iy_0 \in \mathcal{D}_0$ iff for all $t \in \mathbb{R}$

$$(2\theta - 1)\,t^2 + 2(\theta - 1)y_0\,t - (2x_0 + x_0^2 + y_0^2) \ge 0\,,$$

which is for $\theta > \frac{1}{2}$ equivalent with

$$\theta^2 y_0^2 + (2\theta - 1)(1 + x_0)^2 \le 2\theta - 1\,.$$

Plots of these ellipse-shaped regions are given in Figure 4.1. If $\theta = 1$ we recover the stability region of the explicit Euler method, but if we decrease θ

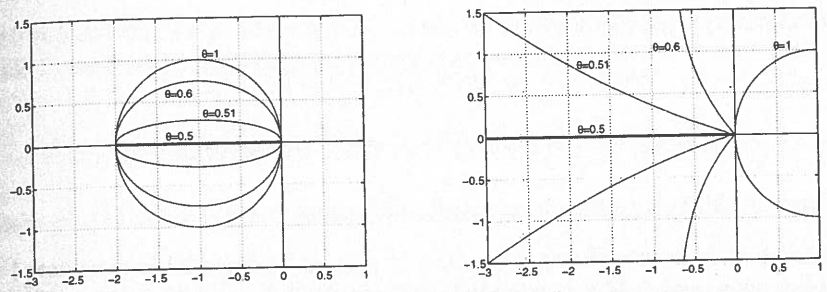

Fig. 4.1. Boundaries of regions $\mathcal{D}_0$ (left) and $\mathcal{D}_1$ (right) for the IMEX-θ method (4.2) with $\theta = 0.5, 0.51, 0.6$ and 1.

the set starts to shrink, and for $\theta = \frac{1}{2}$ it even reduces to the negative line segment $[-2, 0]$.

Alternatively, one can insist on using the full stability region of the explicit method, $\mathcal{S}_0 = \{z_0 : |1 + z_0| \le 1\}$, but then z_1 has to be restricted to the set

$$\mathcal{D}_1 = \{z_1 \in \mathbb{C} : \text{the IMEX scheme is stable for any } z_0 \in \mathcal{S}_0\}\,. \tag{4.6}$$

Then, for all $r_0 = 1+z_0$ in the closed unit disk we should have $|r_0+(1-\theta)z_1| \le |1-\theta z_1|$. The left-hand side is largest if r_0 has the same argument as $(1-\theta)z_1$ and lies on the unit circle. Thus we find that $z_1 \in \mathcal{D}_1$ iff

$$1 + |(1-\theta)\, z_1| \;\le\; |1 - \theta z_1|\,.$$

See again Figure 4.1 for an illustration. Note that it is only for $\theta = 1$ that we get the stability region of the implicit θ-method and that the set $\mathcal{D}_1$ equals the non-positive real line $\mathbb{R}^-$ if $\theta = \frac{1}{2}$.

Method (4.2) with $\theta = 1$ could be viewed as a time splitting method where we first solve $v'(t) = F_0(t, v(t))$ on $[t_n, t_{n+1}]$ with forward Euler and then $v'(t) = F_1(t, v(t))$ with backward Euler. This explains the favourable stability results with $\theta = 1$. However, the structure of the truncation error is very different from the time splitting methods where intermediate results are inconsistent with the full equation. This is due to interference of the first-order splitting error with the first-order Euler errors.

Note that if $\theta > \frac{1}{2}$ the implicit θ-method is *strongly* A-stable (that is, A-stable with damping at ∞), whereas with $\theta = \frac{1}{2}$ the method is 'just' A-stable. Apparently, using a strongly A-stable implicit method gives better stability properties within the IMEX formula (4.2). On the other hand, the above criteria with the sets $\mathcal{D}_0$ and $\mathcal{D}_1$ are rather strict. If we confine z_1 to the negative real line, for example, then also $\theta = \frac{1}{2}$ gives stability for all $|1 + z_0| \le 1$. Therefore, the IMEX-$\theta = \frac{1}{2}$ method should not be discarded, but extra care should be given to stability when applying it.

Finally we note that if F_0 is a genuinely non-stiff term, then we may assume that $|z_0| \le L\tau$ with moderate constant L, in which case we can apply

a simple perturbation argument,

$$|R(z_0,z_1)| \;\le\; |R(0,z_1)| + |1-\theta z_1|^{-1}|z_0|\,,$$

$$|R(z_0,z_1)^n| \;\le\; \big(1+L\tau\big)^n \;\le\; e^{Lt_n} \quad \text{whenever } z_1 \in \mathbb{C}^-,$$

to show stability on finite intervals $[0,T]$ for any $\theta \ge \frac{1}{2}$.

Remark 4.1 In the above the values of λ_0 and λ_1 have been considered as independent, which is a reasonable assumption if F_0 and F_1 act in different directions, for instance with $F_0 \approx a\partial_x$ (horizontal coupling) and $F_1 \approx d\partial_{zz}$ (vertical coupling) or F_1 representing a reaction term (coupling over chemical species). Different results are obtained if there is a dependence between λ_0 and λ_1. Then the implicit treatment of λ_1 can stabilize the process so that even $z_0 \in \mathcal{S}_0$ may no longer be needed. Consider the standard example of the 1D advection-diffusion equation $u_t + au_x = du_{xx}$ with periodicity in space and with second-order spatial discretization. If advection is treated explicitly and diffusion implicitly, then for z_0, z_1 we may take

$$z_0 = -i\nu\sin(2\omega)\,, \qquad z_1 = -4\mu\sin^2(\omega)$$

with $\nu = a\tau/h$, $\mu = d\tau/h^2$ and $0 \le \omega \le \pi$, see Section I.3. A straightforward calculation shows that $|R| \le 1$ iff

$$1 - 8(1-\theta)\mu s + 16(1-\theta)^2\mu^2 s^2 + 4\nu^2 s(1-s) \;\le\; 1 + 8\theta\mu s + 16\theta^2\mu^2 s^2$$

with $s = \sin^2(\omega)$. This holds for all $s \in [0,1]$ iff

$$\nu^2 \le 2\mu \quad \text{and} \quad 2(1-2\theta)\mu \le 1\,.$$

So for any $\theta \ge \frac{1}{2}$ we now just have the condition $\nu^2 \le 2\mu$ for stability, that is $\tau \le 2d/a^2$. ◇

In the following we will discuss several generalizations of the simple θ-method (4.2). Such generalizations are necessary for practical problems since the explicit Euler method is not well suited for advection and also first-order accuracy is often not sufficient.

4.2 IMEX Multistep Methods

With linear multistep methods (II.3.1) the above IMEX approach can be generalized as follows. Consider the fully implicit linear k-step method

$$\sum_{j=0}^{k} \alpha_j w_{n+j} \;=\; \tau \sum_{j=0}^{k} \beta_j \Big(F_0(t_{n+j}, w_{n+j}) + F_1(t_{n+j}, w_{n+j}) \Big) \tag{4.7}$$

with separation of F_0-terms and F_1-terms. We can handle the F_0-terms in an explicit manner by applying the extrapolation formula

$$\varphi(t_{n+k}) = \sum_{j=0}^{k-1} \gamma_j \varphi(t_{n+j}) + \mathcal{O}(\tau^q),$$

where $\varphi(t) = F_0(t, w(t))$. This leads to the k-step IMEX method

$$\sum_{j=0}^{k} \alpha_j w_{n+j} = \tau \sum_{j=0}^{k-1} \beta_j^* F_0(t_{n+j}, w_{n+j}) + \tau \sum_{j=0}^{k} \beta_j F_1(t_{n+j}, w_{n+j}), \qquad (4.8)$$

with new coefficients $\beta_j^* = \beta_j + \beta_k \gamma_j$. Methods of this implicit-explicit multistep type were introduced by Crouzeix (1980) and Varah (1980). The order of consistency of (4.8) is easy to establish.

Theorem 4.2 *Assume the implicit linear multistep method (4.7) has order p and the extrapolation procedure has order q. Then the IMEX method (4.8) has order $r = \min(p, q)$.*

Proof. With $\varphi(t) = F_0(t, w(t))$, the local truncation error can be written as

$$\frac{1}{\tau} \sum_{j=0}^{k} \Big(\alpha_j w(t_{n+j}) - \tau \beta_j w'(t_{n+j}) \Big) + \beta_k \Big(\varphi(t_{n+k}) - \sum_{j=0}^{k-1} \gamma_j \varphi(t_{n+j}) \Big)$$

$$= C\tau^p w^{(p+1)}(t_n) + \mathcal{O}(\tau^{p+1}) + \beta_k C' \tau^q \varphi^{(q)}(t_n) + \mathcal{O}(\tau^{q+1})$$

with constants C, C' determined by the coefficients of the multistep method and the extrapolation procedure. □

In this truncation error only total derivatives of w and φ arise, and therefore the error is not influenced by large Lipschitz constants (negative powers of the mesh width) in F_0 or F_1. This is an important observation since it means absence of order reduction. On the other hand, stability results for the IMEX multistep methods are quite complicated, even for the simple test problem (4.3).

In the remainder of this section we focus on two-step methods and first give three examples of known, popular two-step schemes with $p = q = 2$. In these examples the most advanced time level is taken as t_{n+1}.

Example 4.3 Using the explicit midpoint (Leap-Frog) method for the explicit part and the trapezoidal rule (Crank-Nicolson) for the implicit part yields the popular IMEX-CNLF scheme

$$w_{n+1} - w_{n-1} = 2\tau F_0(t_n, w_n) + \tau F_1(t_{n+1}, w_{n+1}) + \tau F_1(t_{n-1}, w_{n-1}). \qquad (4.9)$$

The stability region $\mathcal{S}_0$ of the explicit method is restricted to the imaginary axis between $-i$ and i, see Example II.3.5. The implicit method is the A-stable trapezoidal rule with step size 2τ. ◇

Example 4.4 A second-order IMEX-BDF scheme can be derived from the implicit two-step backward differentiation formula (II.3.11) and its explicit counterpart (II.3.13). We consider the family of schemes

$$\begin{aligned} \tfrac{3}{2}w_{n+1} - 2w_n + \tfrac{1}{2}w_{n-1} &= 2\tau F_0(t_n, w_n) - \tau F_0(t_{n-1}, w_{n-1}) \\ +\gamma\tau F_1(t_{n+1}, w_{n+1}) &+ 2(1-\gamma)\tau F_1(t_n, w_n) - (1-\gamma)\tau F_1(t_{n-1}, w_{n-1}) \end{aligned} \tag{4.10}$$

with parameter $\gamma \geq 0$. The order is two and the implicit method is A-stable for $\gamma \geq \frac{3}{4}$. With $\gamma = 1$, $F_0 = 0$ we regain the fully implicit BDF2 method.

In applications we will usually take $\gamma = 1$. Higher-order k-step IMEX-BDF type schemes are obtained starting with the fully implicit k-step BDF scheme together with kth order extrapolation for the explicit term, see Crouzeix (1980), Ascher, Ruuth & Wetton (1995). ◇

Example 4.5 The third example is based on a class of second-order two-step Adams methods from Example II.3.2 with a parameter γ,

$$\begin{aligned} w_{n+1} - w_n = \tfrac{3}{2}\tau F_0(t_n, w_n) - \tfrac{1}{2}\tau F_0(t_{n-1}, w_{n-1}) + \gamma\tau F_1(t_{n+1}, w_{n+1}) \\ + \big(\tfrac{3}{2} - 2\gamma\big)\tau F_1(t_n, w_n) + \big(\gamma - \tfrac{1}{2}\big)\tau F_1(t_{n-1}, w_{n-1})\,. \end{aligned} \tag{4.11}$$

The explicit method is the two-step Adams-Bashforth method. The implicit method is A-stable if $\gamma \geq \frac{1}{2}$. If $\gamma = \frac{1}{2}$ the implicit method reduces to the trapezoidal rule. The choice $\gamma = \frac{9}{16}$ was considered by Ascher et al. (1995); this choice yields maximal damping at $z_1 = \infty$. The implicit method with $\gamma = \frac{3}{4}$ was advocated by Nevanlinna & Liniger (1979) with regard to contractivity for scalar problems. ◇

In Figure 4.2 the stability regions $\mathcal{S}_0$ of the explicit methods in (4.10) and (4.11) are plotted together with the regions $\mathcal{D}_0$, defined as in (4.5), where we allow for arbitrary $z_1 \in \mathbb{C}^-$. We see from the figure that $\mathcal{D}_0$ is really smaller than $\mathcal{S}_0$ and if the implicit method is just A-stable, the region $\mathcal{D}_0$ reduces to a line. Formulas for the boundary of $\mathcal{D}_0$ can be found in Frank et al. (1997), where it was also shown that $\mathcal{D}_0 = \mathcal{S}_0$ for the IMEX-CNLF scheme (4.9).

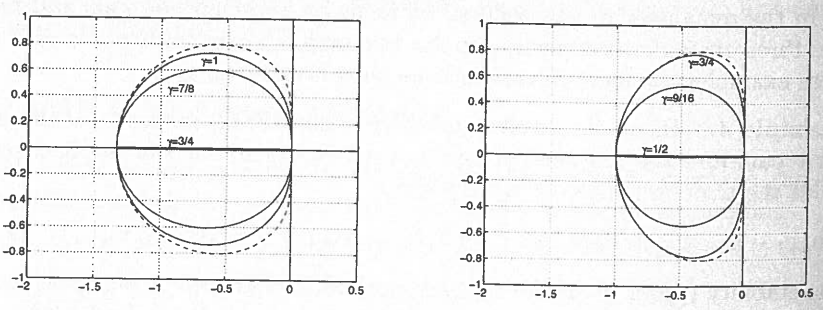

Fig. 4.2. Explicit stability regions $\mathcal{S}_0$ (dashed) and regions $\mathcal{D}_0$ for the IMEX-BDF2 methods (4.10) (left) and two-step IMEX-Adams methods (4.11) (right).

Remark 4.6 Stability in actual applications is determined by specific spatial operators and selected spatial discretizations. Often, the non-stiff part F_0 emanates from advection and the stiff part F_1 from reaction and diffusion terms. We here consider the specific case that λ_0 is associated to advection discretized by the first-order upwind or third-order upwind-biased scheme, while λ_1 may still take on arbitrary values in $\mathbb{C}^-$. For a given IMEX method this leads to a CFL restriction. In a manner similar to Table II.3.1, such restrictions are given in Table 4.1 for the IMEX-BDF2 scheme (4.10) with $\gamma = 1$ and the IMEX-Adams scheme (4.11) with $\gamma = \frac{3}{4}, \frac{9}{16}$ for z_0 in $\mathcal{S}_0$ or $\mathcal{D}_0$. The numbers have been determined experimentally by comparisons of the eigenvalues with the stability sets of Figure 4.2.

	BDF2, $\gamma = 1$		Adams2, $\gamma = \frac{3}{4}$		Adams2, $\gamma = \frac{9}{16}$	
	$\mathcal{S}_0$	$\mathcal{D}_0$	$\mathcal{S}_0$	$\mathcal{D}_0$	$\mathcal{S}_0$	$\mathcal{D}_0$
$z_0 = z_{a,1}$	0.66	0.66	0.50	0.50	0.50	0.50
$z_0 = z_{a,3}$	0.46	0.23	0.58	0.43	0.58	0.16

Table 4.1. CFL restrictions for the IMEX methods (4.10) and (4.11) with first-order upwind $z_{a,1}$ and third-order upwind-biased advection $z_{a,3}$ (as in Table II.3.1).

With regard to applications, the results for the third-order upwind-biased discretization are more important than those for first-order upwind. For the latter discretization the CFL restrictions are the same for $\mathcal{S}_0$ and $\mathcal{D}_0$. With the third-order discretization the requirement of A-stability for the implicit term has a large effect on the allowable Courant number of the IMEX-BDF2 method; even though the region $\mathcal{D}_0$ seems only slightly smaller than $\mathcal{S}_0$ in Figure 4.2, this results in a reduction of the maximal Courant number by approximately one half. The reason for this is that eigenvalues of the third-order scheme are very close to the imaginary axis in the vicinity of the origin. In this respect, among the two-step IMEX schemes considered here, the Adams method (4.11) with $\gamma = \frac{3}{4}$ gives the best results.

Central advection discretization of even order leads to purely imaginary eigenvalues. Therefore, among the two-step methods considered here only the Crank-Nicolson Leap-Frog method (4.9) will be stable. For the other two-step methods some upwinding or diffusion is necessary. The Leap-Frog method cannot be used with upwinding, of course. ◇

Although A-stability is a valuable property, in many practical situations one can settle for less, such as $A(\alpha)$-stability. Some sufficient analytical results for stability with arbitrary $z_0 \in \mathcal{S}_0$ and with z_1 in a wedge $\mathcal{W}_\alpha$, i.e., $|\arg(-z_1)| \leq \alpha$, were obtained by Frank et al. (1997). Some pictures of the sets $\mathcal{D}_1$, defined as in (4.6), are displayed in Figure 4.3; these pictures were

obtained numerically. By zooming in on the origin one can establish an experimental bound of the admissible angles α for stability with arbitrary $z_0 \in \mathcal{S}_0$ and $z_1 \in \mathcal{W}_\alpha$. For the IMEX-BDF2 method (4.10) with $\gamma = 1$ it was found that $\alpha \approx 0.32\pi$. For the IMEX-Adams method (4.11) with $\gamma = \frac{3}{4}$ and $\gamma = \frac{9}{16}$ the experimental bound was found to be $\alpha \approx 0.30\pi$ and $\alpha \approx 0.14\pi$, respectively. In Frank et al. (1997) it was also shown that for the IMEX-CNLF scheme (4.9) the A-stability property is retained.

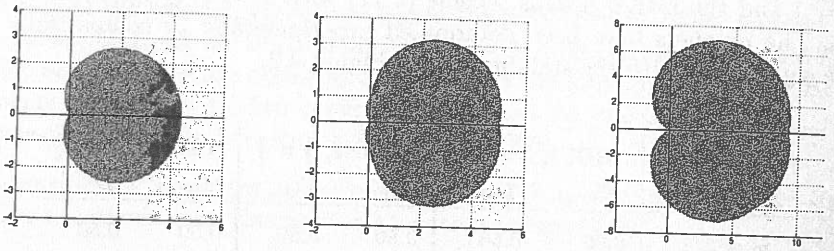

Fig. 4.3. Exterior of shaded region, $\mathcal{D}_1$: stability for z_1 with arbitrary $z_0 \in \mathcal{S}_0$ for the IMEX-BDF2 method with $\gamma = 1$ (left) and the two-step IMEX-Adams method with $\gamma = \frac{3}{4}$ (middle) and $\gamma = \frac{9}{16}$ (right).

Above we followed a standard ODE stability analysis in the sense that the eigenvalues λ_0 and λ_1 were allowed to take on arbitrary complex values in certain regions in the complex plane. In actual applications these regions are determined by specific spatial discretizations. Often, the non-stiff part F_0 emanates from advection and the stiff part F_1 from reaction-diffusion terms. For example, for atmospheric transport-chemistry models a useful test model is the system $u_t + a_1 u_x + a_2 u_y = \epsilon u_{zz} + f(u)$ where u is a vector of chemical concentrations, $a_1 u_x + a_2 u_y$ models advection in a horizontal wind field, ϵu_{zz} models a vertical diffusion process, which includes parameterized turbulence, and $f(u)$ is a set of atmospheric stiff chemical reactions. Application of IMEX schemes to atmospheric problems has been investigated in Verwer et al. (1996).

Results for 1D linear advection-diffusion equations $u_t + a u_x = d u_{xx}$ with constant coefficients, based on Fourier decompositions as in Remark 4.1, can be found in Varah (1980), Ascher et al. (1995). Sufficient conditions for multi-dimensional problems with constant coefficients are found in Wesseling (2001). More general stability results, valid for noncommuting operators, are given in Crouzeix (1980). Generalizations to problems that are nonlinear in F_0 can be found in Akrivis et al. (1999).

Remark 4.7 Multistep methods for splittings (3.5) with s implicit terms can be derived by using the stabilizing corrections idea, where one starts with an explicit scheme and then adds implicitness as corrections. For $s = 1$ this leads to IMEX methods. For $s \geq 2$, however, this seems to give quite