15. The Method of Lumped Masses

In this chapter we shall consider a modification of the standard Galerkin method using piecewise linear trial functions, the so-called method of lumped masses. In this method the mass matrix is replaced by a diagonal matrix with the row sums of the original mass matrix as its diagonal elements. This can also be interpreted as using a quadrature rule for the corresponding L_2 inner product.

We consider the simple initial-boundary value problem

$$\begin{aligned} u_t - \Delta u = f \quad \text{in } \Omega, \quad t > 0, \\ u = 0 \quad \text{on } \partial\Omega,\ t\ > 0, \quad \text{with } u(\cdot, 0) = v \quad \text{in } \Omega, \end{aligned}$$

where again for simplicity Ω is a smooth convex domain in the plane.

Let $S_h \subset H_0^1 = H_0^1(\Omega)$ consist of continuous, piecewise linear functions on a quasiuniform family of triangulations $\mathcal{T}_h = \{\tau\}$ of Ω with its boundary vertices on $\partial\Omega$ and which vanish outside the polygonal domain Ω_h determined by $\mathcal{T}_h$. Let $\{P_j\}_{j=1}^{N_h}$ denote the interior vertices of $\mathcal{T}_h$ and let $\{\Phi_j\}_{j=1}^{N_h}$ be the standard basis for S_h consisting of the pyramid functions defined by $\Phi_j(P_k) = \delta_{jk}$.

Recall that the basic semidiscrete Galerkin method is to find $u_h : [0, \infty) \to S_h$ such that

$$(u_{h,t}, \chi) + (\nabla u_h, \nabla \chi) = (f, \chi), \quad \forall \chi \in S_h,\ t > 0, \quad u_h(0) = v_h, \tag{15.1}$$

where v_h is some approximation of v in S_h. Recall also that this method may be written in matrix form as

$$\mathcal{B}\alpha'(t) + \mathcal{A}\alpha(t) = \widetilde{f}(t), \quad \text{for } t > 0, \quad \text{with } \alpha(0) = \gamma, \tag{15.2}$$

where $\mathcal{B} = (b_{jk})$ and $\mathcal{A} = (a_{jk})$ are the mass and stiffness matrices with elements $b_{jk} = (\Phi_j, \Phi_k)$ and $a_{jk} = (\nabla\Phi_j, \nabla\Phi_k)$, respectively, where $\alpha_j(t)$ and γ_j are the coefficients of $u_h(t)$ and v_h with respect to $\{\Phi_j\}_{j=1}^{N_h}$ and where $\widetilde{f}$ is the vector with components (f, Φ_k).

The lumped mass method consists in replacing the mass matrix $\mathcal{B}$ in (15.2) by the diagonal matrix $\bar{\mathcal{B}}$ obtained by taking for its diagonal elements the numbers $\bar{b}_{jj} = \sum_{k=1}^{N_h} b_{jk}$, i.e., by lumping all masses in one row into the diagonal entry. This makes the inversion of the matrix coefficient of $\alpha'(t)$ a triviality.

We shall thus study the matrix problem

$$\bar{\mathcal{B}}\alpha'(t) + \mathcal{A}\alpha(t) = \widetilde{f}(t), \quad \text{for } t > 0, \quad \text{with } \alpha(0) = \gamma. \tag{15.3}$$

We shall now describe two alternative interpretations of this procedure, and then use the first of these to show some error estimates for it.

Our first interpretation will be to think of (15.3) as being obtained by evaluating the first term in (15.1) by numerical quadrature. Let τ be a triangle of the triangulation $\mathcal{T}_h$, let $P_{\tau,j}$, $j = 1, 2, 3$, be its vertices, and consider the quadrature formula

$$Q_{\tau,h}(f) = \tfrac{1}{3} \text{ area } (\tau) \sum_{j=1}^{3} f(P_{\tau,j}) \approx \int_\tau f\,dx. \tag{15.4}$$

We may then define an approximation of the inner product in S_h by

$$(\psi, \chi)_h = \sum_{\tau \in \mathcal{T}_h} Q_{\tau,h}(\psi\chi). \tag{15.5}$$

We claim now that the lumped mass method defined by (15.3) above is equivalent to

$$(u_{h,t}, \chi)_h + (\nabla u_h, \nabla \chi) = (f, \chi), \quad \forall \chi \in S_h,\ t > 0, \quad u_h(0) = v_h. \tag{15.6}$$

In fact, setting $u_h(t) = \sum_{j=1}^{N_h} \alpha_j(t)\Phi_j$ this system may be written

$$\sum_{j=1}^{N_h} \alpha_j'(t)(\Phi_j, \Phi_k)_h + \sum_{j=1}^{N_h} \alpha_j(t)(\nabla\Phi_j, \nabla\Phi_k) = (f, \Phi_k), \quad k = 1, \ldots, N_h,$$

and to show the equivalence it remains only to observe that $(\Phi_j, \Phi_k)_h = 0$ for $j \neq k$, as $\Phi_j(x)\Phi_k(x)$ vanishes at all vertices of $\mathcal{T}_h$, and to show that

$$\|\Phi_j\|_h^2 = (\Phi_j, \Phi_j)_h = \sum_{k=1}^{N_h} (\Phi_j, \Phi_k). \tag{15.7}$$

To prove this latter fact, note that (Φ_j, Φ_k) is only non-zero for $j \neq k$ if P_j and P_k are neighbors, and observe that in such a case, if τ is a triangle with P_j and P_k as vertices, simple calculations, for instance after transformation to a reference triangle, show that

$$\int_\tau \Phi_j\Phi_k\,dx = \tfrac{1}{12} \text{ area } (\tau) \quad \text{and} \quad \int_\tau \Phi_j^2\,dx = \tfrac{1}{6} \text{ area } (\tau).$$

It follows, since for each pair of neighbors P_j, P_k there are two such triangles τ, that with D_j the union of the triangles which have P_j as a vertex,

$$\sum_{k=1}^{N_h} (\Phi_j, \Phi_k) = \tfrac{1}{3} \text{ area } (D_j).$$

Since clearly

$$\|\Phi_j\|_h^2 = \sum_\tau Q_{\tau,h}(\Phi_j^2) = \tfrac{1}{3} \text{ area } (D_j),$$

this completes the proof of (15.7).

We now turn to the other formulation of the method under consideration. Let again τ be a triangle of the triangulation and P_j one of its vertices. Now draw the straight lines connecting each vertex of τ to the midpoint of the opposite side of τ. These straight lines intersect at the barycenter of τ and divide τ into six triangles of equal area. Let $B_{j,\tau}$ be the union of the two of these that have P_j as a vertex. Clearly, then, the area of $B_{j,\tau}$ is a third of that of τ. For each interior vertex P_j, let B_j be the union of the $B_{j,\tau}$ for which τ has P_j as a vertex.

Now let $\bar{S}_h$ denote the functions which are constant on each B_j and vanish outside the union of the B_j. We note that the elements $\bar{\chi}$ of $\bar{S}_h$ are uniquely defined by the values at the vertices P_j and we may write

$$\bar{\chi}(x) = \sum_{j=1}^{N_h} \bar{\chi}(P_j)\bar{\Phi}_j(x),$$

where $\bar{\Phi}_j = 1$ on B_j and vanishes elsewhere. Since the functions of S_h are also uniquely determined by their values at the P_j there is a one-to-one correspondence between the functions of S_h and those of $\bar{S}_h$, and for χ in S_h we denote by $\bar{\chi}$ the associated function in $\bar{S}_h$ which agrees with χ at the P_j.

With this notation the semidiscrete equation (15.3) or (15.6) may also be formulated as

$$(\bar{u}_{h,t}, \bar{\chi}) + (\nabla u_h, \nabla \chi) = (f, \chi), \ \forall \chi \in S_h, \ t > 0, \quad u_h(0) = v_h.$$

In fact, this follows similarly to above if we observe that trivially $(\bar{\Phi}_j, \bar{\Phi}_k) = 0$ for $j \neq k$ and that $\|\bar{\Phi}_j\|^2 = \text{ area } (B_j) = \text{ area } (D_j)/3 = \|\Phi_j\|_h^2$.

One may think of this latter formulation as being obtained by reducing the H^1 regularity requirements for the functions in S_h in the first term of (15.1), where they are not needed for the products to make sense. This latter approach was taken in [203] and [49].

We now turn to the error analysis and return to the formulation (15.6). We begin with the following lemma.

Lemma 15.1 *Let $\varepsilon_h(v, w) = (v, w)_h - (v, w)$ denote the quadrature error in* (15.5). *We then have*

$$|\varepsilon_h(\psi, \chi)| \le Ch^2 \|\nabla \psi\| \, \|\nabla \chi\|, \quad \textit{for } \psi, \chi \in S_h.$$

Proof. Since the quadrature formula (15.4) is exact for f linear we have, by transformation to a fixed reference triangle τ_0 and using the Bramble-Hilbert lemma and the Sobolev inequality $\|f\|_{L_\infty(\tau_0)} \le C\|f\|_{W_1^2(\tau_0)}$, that

$$|Q_{\tau,h}(f) - \int_\tau f\,dx| \le Ch^2 \sum_{|\alpha|=2} \|D^\alpha f\|_{L_1(\tau)}.$$

After application to $f = \psi\chi$ this implies, since both ψ and χ are linear in τ, that

$$|Q_{\tau,h}(\psi\chi) - \int_\tau \psi\chi\,dx| \le Ch^2\|\nabla\psi\|_{L_2(\tau)}\|\nabla\chi\|_{L_2(\tau)}.$$

Using the Cauchy-Schwarz inequality we conclude that

$$|\varepsilon_h(\psi,\chi)| \le Ch^2 \sum_{\tau\in\mathcal{T}_h} \|\nabla\psi\|_{L_2(\tau)}\|\nabla\chi\|_{L_2(\tau)} \le Ch^2\|\nabla\psi\|\,\|\nabla\chi\|,$$

which is the desired estimate. □

We shall now show the following error estimate:

Theorem 15.1 *We have for the error in the semidiscrete lumped mass method* (15.6), *for* $t \ge 0$,

$$\|u_h(t) - u(t)\| \le C\|v_h - v\| + Ch^2\Big(\|v\|_2 + \|u(t)\|_2 + \big(\int_0^t \|u_t\|_2^2\,ds\big)^{1/2}\Big).$$

Proof. We write, with R_h the standard Ritz projection, $u_h - u = (u_h - R_hu) + (R_hu - u) = \theta + \rho$, and $\rho(t)$ is bounded in the desired way. In order to estimate θ, we write

$$\begin{aligned}
&(\theta_t,\chi)_h + (\nabla\theta,\nabla\chi)\\
&\quad = (u_{h,t},\chi)_h + (\nabla u_h,\nabla\chi) - (R_hu_t,\chi)_h - (\nabla R_hu,\nabla\chi)\\
&\quad = (f,\chi) - (R_hu_t,\chi)_h - (\nabla u,\nabla\chi) = (u_t,\chi) - (R_hu_t,\chi)_h\\
&\quad = -(\rho_t,\chi) - \varepsilon_h(R_hu_t,\chi).
\end{aligned} \tag{15.8}$$

Setting $\chi = \theta$ we obtain

$$\tfrac12 \frac{d}{dt}\|\theta\|_h^2 + \|\nabla\theta\|^2 = -(\rho_t,\theta) - \varepsilon_h(R_hu_t,\theta). \tag{15.9}$$

Here we have at once

$$|(\rho_t,\theta)| \le \|u_t - R_hu_t\|\,\|\theta\| \le Ch^2\|u_t\|_2\|\theta\| \le Ch^2\|u_t\|_2\|\nabla\theta\|,$$

and, using Lemma 15.1,

$$|\varepsilon_h(R_hu_t,\theta)| \le Ch^2\|\nabla R_hu_t\|\,\|\nabla\theta\| \le Ch^2\|u_t\|_2\|\nabla\theta\|.$$

It follows thus that

$$\tfrac{1}{2}\frac{d}{dt}\|\theta\|_h^2 + \|\nabla\theta\|^2 \le Ch^2\|u_t\|_2\|\nabla\theta\| \le \|\nabla\theta\|^2 + Ch^4\|u_t\|_2^2,$$

from which we infer

$$\|\theta(t)\|_h^2 \le \|\theta(0)\|_h^2 + Ch^4\int_0^t \|u_t\|_2^2\,ds.$$

We now note that $\|\cdot\|_h$ and $\|\cdot\|$ are equivalent norms on S_h, uniformly in h (this follows easily by considering each triangle separately), and that hence

$$\|\theta(t)\| \le C\|\theta(0)\| + Ch^2\Big(\int_0^t \|u_t\|_2^2\,ds\Big)^{1/2}.$$

Here $\|\theta(0)\| = \|v_h - R_h v\| \le \|v_h - v\| + Ch^2\|v\|_2$, whence $\theta(t)$ is bounded as desired. The proof is complete. □

We now turn to an estimate for the gradient.

Theorem 15.2 *We have for the error in the semidiscrete method* (15.6)*, for* $t \ge 0$,

$$\|\nabla(u_h - u)(t)\| \le \|\nabla(v_h - v)\| + Ch\Big(\|v\|_2 + \|u(t)\|_2 + \Big(\int_0^t \|\nabla u_t\|^2\,ds\Big)^{1/2}\Big).$$

Proof. We now set $\chi = \theta_t$ in the equation (15.8) for θ to obtain

$$\|\theta_t\|_h^2 + \tfrac{1}{2}\frac{d}{dt}\|\nabla\theta\|^2 = -(\rho_t, \theta_t) - \varepsilon_h(R_h u_t, \theta_t). \tag{15.10}$$

Here, as in the proof of Theorem 15.1,

$$|(\rho_t, \theta_t)| \le \|u_t - R_h u_t\|\,\|\theta_t\| \le Ch\|\nabla u_t\|\,\|\theta_t\|.$$

Further, by Lemma 15.1,

$$|\varepsilon_h(R_h u_t, \theta_t)| \le Ch^2\|\nabla R_h u_t\|\,\|\nabla\theta_t\| \le Ch\|\nabla u_t\|\,\|\theta_t\|,$$

where in the last step we have applied the inverse estimate (1.12). (The use of the inverse estimate may be avoided by a slight modification of Lemma 15.1.) Using again the equivalence between the norms $\|\cdot\|_h$ and $\|\cdot\|$ on S_h we conclude

$$\|\theta_t\|_h^2 + \tfrac{1}{2}\frac{d}{dt}\|\nabla\theta\|^2 \le Ch\|\nabla u_t\|\,\|\theta_t\|_h \le \|\theta_t\|_h^2 + Ch^2\|\nabla u_t\|^2,$$

so that, after integration,

$$\begin{aligned}\|\nabla\theta(t)\| &\le \|\nabla\theta(0)\| + Ch\Big(\int_0^t \|\nabla u_t\|^2\,ds\Big)^{1/2}\\ &\le \|\nabla(v_h - v)\| + Ch\Big(\|v\|_2 + \Big(\int_0^t \|\nabla u_t\|^2\,ds\Big)^{1/2}\Big).\end{aligned}$$

Together with the standard estimate for $\nabla\rho(t)$ this completes the proof. □

This demonstration does not immediately yield the superconvergent order $O(h^2)$ estimate for $\nabla\theta$ which holds for the standard Galerkin method. However, as is shown in the following lemma, a slight modification of the proof shows such a result.

Lemma 15.2 *For each $\bar{t} > 0$ there is a constant $C = C(\bar{t})$ such that for $\theta = u_h - R_h u$ and $0 \le t \le \bar{t}$,*

$$\|\nabla\theta(t)\| \le \|\nabla\theta(0)\| + Ch^2\Big(\|u_t(t)\|_2 + \big(\int_0^t (\|u_t\|_2^2 + \|u_{tt}\|_1^2)\,ds\big)^{1/2}\Big).$$

Proof. It suffices to consider the case $v_h = R_h v$, or $\theta(0) = 0$. For the solution $\widetilde{u}_h$ of the homogeneous equation with initial data $\widetilde{u}_h(0) = v_h - R_h v = \theta(0)$ satisfies

$$\|\widetilde{u}_{h,t}\|_h^2 + \tfrac{1}{2}\frac{d}{dt}\|\nabla\widetilde{u}_h\|^2 = 0,$$

and hence

$$\|\nabla\widetilde{u}_h(t)\|^2 \le \|\nabla\widetilde{u}_h(0)\|^2 = \|\nabla\theta(0)\|^2.$$

We have as before (15.10), which we now write in the form

$$\|\theta_t\|_h^2 + \tfrac{1}{2}\frac{d}{dt}\|\nabla\theta\|^2 = -(\rho_t, \theta_t) - \frac{d}{dt}\varepsilon_h(R_h u_t, \theta) + \varepsilon_h(R_h u_{tt}, \theta). \tag{15.11}$$

Here

$$|(\rho_t, \theta_t)| \le \|\rho_t\|\,\|\theta_t\| \le Ch^2\|u_t\|_2\|\theta_t\|_h \le Ch^4\|u_t\|_2^2 + \|\theta_t\|_h^2.$$

Further, by Lemma 15.1,

$$|\varepsilon_h(R_h u_t, \theta)| \le Ch^2\|\nabla R_h u_t\|\,\|\nabla\theta\| \le Ch^4\|u_t\|_1^2 + \tfrac{1}{4}\|\nabla\theta\|^2,$$

and similarly with u_t replaced by u_{tt}. By integration of (15.11) we therefore obtain, since $\theta(0) = 0$,

$$\|\nabla\theta(t)\|^2 \le Ch^4\Big(\|u_t(t)\|_1^2 + \int_0^t (\|u_t\|_2^2 + \|u_{tt}\|_1^2)\,ds\Big) + \int_0^t \|\nabla\theta\|^2\,ds.$$

The result now follows by Gronwall's lemma. □

As one application of the lemma we shall prove the following maximum-norm error estimate:

Theorem 15.3 *Let v_h be chosen so that $\|\nabla(v_h - R_h v)\| \le Ch^2$. Then under the appropriate regularity assumptions we have for the error in* (15.6)

$$\|u_h(t) - u(t)\|_{L_\infty} \le C(\bar{t}; u)h^2\ell_h, \quad \text{where } \ell_h = \max(1, \log(1/h)), \quad \text{for } t \le \bar{t}.$$

Proof. We recall that since the triangulation is quasiuniform we may apply the "almost" Sobolev inequality of Lemma 6.4 together with Lemma 15.2 to obtain

$$\|\theta(t)\|_{L_\infty} \le C(\bar{t};u)h^2\ell_h^{1/2}.$$

In view of the maximum-norm error estimate of Theorem 1.4 for the elliptic problem, this shows the result. □

We observe that because of the use of quadrature, our above error analyses of Theorem 15.1 and Lemma 15.2 require more regularity of the solution than was the case for the standard Galerkin method. For the homogeneous equation, for instance, Theorem 15.1 shows by standard calculations using the definition of the norm in $\dot{H}^s = \dot{H}^s(\Omega)$ (cf. Chapter 3) that, for $v_h = R_h v$, say,

$$\|u_h(t) - u(t)\| \le Ch^2|v|_3, \quad \text{for } v \in \dot{H}^3,$$

and Lemma 15.2 shows similarly

$$\|\nabla\theta(t)\| \le Ch^2|v|_4, \quad \text{for } v \in \dot{H}^4.$$

In addition to smoothness these estimates require $v = \Delta v = 0$ on $\partial\Omega$. We shall demonstrate now how at least the latter boundary condition may be removed for t positive, by using our previous techniques for nonsmooth data error estimates.

Lemma 15.3 *Consider the homogeneous equation* ($f = 0$) *and let* $\theta = u_h - R_h u$. *Then for each* $\bar{t} > 0$ *there is a constant* C *such that if* $\theta(0) = 0$ *then for* $0 < t \le \bar{t}$ *and* $v \in \dot{H}^2$,

$$\|\theta(t)\| \le Ch^2 t^{-1/2}|v|_2 \quad \textit{and} \quad \|\nabla\theta(t)\| \le Ch^2 t^{-1}|v|_2.$$

Proof. Multiplying (15.9) by t we have

$$\tfrac{1}{2}\frac{d}{dt}(t\|\theta\|_h^2) + t\|\nabla\theta\|^2 = -t(\rho_t,\theta) - t\varepsilon_h(R_h u_t,\theta) + \tfrac{1}{2}\|\theta\|_h^2.$$

Hence by integration and routine estimates

$$\begin{aligned} t\|\theta\|_h^2 + \int_0^t s\|\nabla\theta\|^2\,ds &\le Ch^4\int_0^t (s^2\|u_t\|_2^2 + s\|u_t\|_1^2)\,ds \\ +C\int_0^t \|\theta\|_h^2\,ds &\le Ch^4\|v\|_1^2 + C\int_0^t \|\theta\|_h^2\,ds. \end{aligned} \tag{15.12}$$

In order to estimate the latter integral we set $\widetilde{\theta}(t) = \int_0^t \theta(s)\,ds$ and integrate the error equation (15.8) from 0 to t to obtain

$$(\theta,\chi)_h + (\nabla\widetilde{\theta},\nabla\chi) = (\rho(0)-\rho(t),\chi) - \varepsilon_h(R_h(u(t)-v),\chi), \quad \forall\chi\in S_h.$$

Setting $\chi = \theta = \widetilde{\theta}_t$ this yields

$$\|\theta\|_h^2 + \tfrac{1}{2}\frac{d}{dt}\|\nabla\widetilde{\theta}\|^2$$
$$= (\rho(0)-\rho(t),\theta) - \frac{d}{dt}\varepsilon_h(R_h(u(t)-v),\widetilde{\theta}) + \varepsilon_h(R_h u_t,\widetilde{\theta}),$$

and hence, by obvious estimates,

$$\int_0^t \|\theta\|_h^2\,ds + \|\nabla\widetilde{\theta}\|^2 \le Ch^4\int_0^t (\|u(s)\|_2 + \|v\|_2)^2\,ds$$
$$+ Ch^4\|\nabla R_h(u(t)-v)\|^2 + Ch^4\int_0^t \|\nabla R_h u_t\|^2\,ds + \int_0^t \|\nabla\widetilde{\theta}\|^2\,ds,$$

so that, using also Gronwall's lemma, for $t \le \bar{t}$,

$$\int_0^t \|\theta\|_h^2\,ds \le Ch^4\big(\|v\|_2^2 + \int_0^t \|u\|_3^2\,ds\big) \le Ch^4|v|_2^2. \tag{15.13}$$

Together with (15.12) this proves the first estimate of the lemma.

In order to bound $\nabla\theta$ we multiply (15.11) by t^2 and obtain easily

$$\tfrac{1}{2}\frac{d}{dt}(t^2\|\nabla\theta\|^2) \le -\frac{d}{dt}(t^2\varepsilon_h(R_h u_t,\theta)) + Ct^2\|\rho_t\|^2$$
$$+ t^2\varepsilon_h(R_h u_{tt},\theta) + 2t\varepsilon_h(R_h u_t,\theta) + t\|\nabla\theta\|^2,$$

or

$$t^2\|\nabla\theta(t)\|^2 \le Ch^4\big(t^2\|\nabla u_t(t)\|^2 + \int_0^t (s^3\|u_{tt}\|_1^2 + s^2\|u_t\|_2^2 + s\|u_t\|_1^2)\,ds\big)$$
$$+ C\int_0^t s\|\nabla\theta\|^2\,ds \le Ch^4 t\|v\|_2^2 + C\int_0^t s\|\nabla\theta\|^2\,ds.$$

Hence we have using (15.12) and (15.13), since $t \le \bar{t}$,

$$t^2\|\nabla\theta(t)\|^2 \le Ch^4|v|_2^2,$$

which completes the proof. □

It is obvious how Lemma 15.3 may be combined with our different estimates for the error ρ in the elliptic projection to yield L_2 and L_∞ norm bounds for the error in the homogeneous semidiscrete equation with initial data $v \in \dot{H}^2$. We shall not insist on the details.

The method of lumped masses may, of course, be used in fully discrete methods. With $\bar{\partial}$ as usual denoting the backward difference quotient and $0 \le \kappa \le 1$ one could, for instance, consider the method defining $U^n = U_h^n \in S_h$ by,

$$\begin{aligned}&(\bar{\partial}U^n,\chi)_h + \kappa(\nabla U^n,\nabla\chi) + (1-\kappa)(\nabla U^{n-1},\nabla\chi)\\ &\quad= (f(t_{n-1}+\kappa k),\chi), \quad \forall\,\chi\in S_h,\ n\ge 1, \quad \text{with } U^0 = v_h,\end{aligned} \tag{15.14}$$

or in matrix form, with α^n the vector of the components of U^n with respect to the basis $\{\Phi_j\}_{j=1}^{N_h}$, and $F^{n-1+\kappa}$ that with components $(f(t_{n-1}+k\kappa), \Phi_j)$,

$$\bar{\mathcal{B}}\bar{\partial}\alpha^n + \kappa\mathcal{A}\alpha^n + (1-\kappa)\mathcal{A}\alpha^{n-1} = F^{n-1+\kappa},$$

or since $\bar{\mathcal{B}} + \kappa k\mathcal{A}$ is obviously positive definite,

$$\alpha^n = (\bar{\mathcal{B}} + \kappa k\mathcal{A})^{-1}(\bar{\mathcal{B}} - (1-\kappa)k\mathcal{A})\alpha^{n-1} + (\bar{\mathcal{B}} + \kappa k\mathcal{A})^{-1}kF^{n-1+\kappa}.$$

The backward Euler method corresponds to $\kappa = 1$, the Crank-Nicolson method to $\kappa = \frac{1}{2}$, and for $\kappa = 0$ we now have a method which is purely explicit since $\bar{\mathcal{B}}$ is diagonal.

As an example, let us briefly analyze the backward Euler method and show the following.

Theorem 15.4 *We have for the backward Euler Galerkin method* (15.14) *with $\kappa = 1$, for $t_n \geq 0$,*

$$\begin{aligned}&\|U^n - u(t_n)\| \leq C\|v_h - v\| \\ &+ Ch^2\Big(\|v\|_2 + \|u(t_n)\|_2 + \big(\int_0^{t_n} \|u_t\|_2^2\,ds\big)^{1/2}\Big) + Ck\big(\int_0^{t_n} \|u_{tt}\|^2\,ds\big)^{1/2}.\end{aligned}$$

Proof. Writing as usual $U^n - u^n = \theta^n + \rho^n$, we need only bound θ^n. We have

$$\begin{aligned}&(\bar{\partial}\theta^n, \chi)_h + (\nabla\theta^n, \nabla\chi) \\ &= (\bar{\partial}U^n, \chi)_h + (\nabla U^n, \nabla\chi) - (\bar{\partial}R_h u^n, \chi)_h - (\nabla R_n u^n, \nabla\chi) \\ &= (f^n, \chi) - (\bar{\partial}R_h u^n, \chi)_h - (\nabla u^n, \nabla\chi) = (u_t^n, \chi) - (\bar{\partial}R_h u^n, \chi)_h.\end{aligned}$$

Choosing $\chi = \theta^n$ we find after some manipulation

$$\begin{aligned}&\frac{1}{2k}(\|\theta^n\|_h^2 - \|\theta^{n-1}\|_h^2) + \frac{1}{2k}\|\theta^n - \theta^{n-1}\|_h^2 + \|\nabla\theta^n\|^2 \\ &= (u_t^n - \bar{\partial}u^n, \theta^n) + (\bar{\partial}u^n - \bar{\partial}R_h u^n, \theta^n) - \varepsilon_h(\bar{\partial}R_h u^n, \theta^n) = R_1 + R_2 + R_3.\end{aligned}$$

We have for the contribution of the discretization in time

$$\begin{aligned}|R_1| &\leq \|u_t^n - \bar{\partial}u^n\|\,\|\theta^n\| \leq C\int_{t_{n-1}}^{t_n} \|u_{tt}\|\,ds\,\|\nabla\theta^n\| \\ &\leq Ck^{1/2}\Big(\int_{t_{n-1}}^{t_n} \|u_{tt}\|^2\,ds\Big)^{1/2}\|\nabla\theta^n\|.\end{aligned}$$

Further

$$\begin{aligned}&|R_2| \leq \|(I - R_h)\bar{\partial}u^n\|\,\|\theta^n\| \leq Ch^2\|\bar{\partial}u^n\|_2\|\theta^n\| \\ &\leq Ch^2k^{-1}\int_{t_{n-1}}^{t_n} \|u_t\|_2\,ds\|\nabla\theta^n\| \leq Ch^2k^{-1/2}\Big(\int_{t_{n-1}}^{t_n} \|u_t\|_2^2\,ds\Big)^{1/2}\|\nabla\theta^n\|,\end{aligned}$$

and finally, using again Lemma 15.1,

$$|R_3| \le Ch^2\|\nabla R_h\bar\partial u^n\|\,\|\nabla\theta^n\| \le Ch^2\|\bar\partial\nabla u^n\|\,\|\nabla\theta^n\|$$
$$\le Ch^2k^{-1}\int_{t_{n-1}}^{t_n}\|\nabla u_t\|\,ds\,\|\nabla\theta^n\| \le Ch^2k^{-1/2}\Big(\int_{t_{n-1}}^{t_n}\|u_t\|_2^2\,ds\Big)^{1/2}\|\nabla\theta^n\|.$$

Altogether we conclude, after a kickback of $k\|\nabla\theta^n\|^2$,

$$\|\theta^n\|_h^2 \le \|\theta^{n-1}\|_h^2 + Ch^4\int_{t_{n-1}}^{t_n}\|u_t\|_2^2\,ds + Ck^2\int_{t_{n-1}}^{t_n}\|u_{tt}\|^2\,ds,$$

and hence

$$\|\theta^n\|_h^2 \le \|\theta^0\|_h^2 + Ch^4\int_0^{t_n}\|u_t\|_2^2\,ds + Ck^2\int_0^{t_n}\|u_{tt}\|^2\,ds.$$

By the equivalence of $\|\cdot\|_h$ and $\|\cdot\|$ on S_h and the standard estimate for $\|\theta^0\|$ this concludes the proof of the theorem. □

We shall now show that if the triangulations used are of Delaunay type, then there is a maximum-principle associated with the lumped mass method. Recall from the beginning of Chapter 6 that this is not the case for the standard Galerkin method. A triangulation is of Delaunay type if for all edges e of $\mathcal{T}_h$, with α_1 and α_2 the two angles opposite to e in the two triangles τ_1 and τ_2 determined by e, respectively, we have $\alpha_1+\alpha_2 \le \pi$. This condition is satisfied, in particular, if all angles α of $\mathcal{T}_h$ are acute. Note that this does not require $\mathcal{T}_h$ to be quasiuniform.

We consider the homogeneous equation

$$(u_{h,t},\chi)_h + (\nabla u_h,\nabla\chi) = 0, \quad \forall\chi\in S_h,\ t>0, \quad \text{with } u_h(0)=v_h,$$

and we denote by $\bar E_h(t): S_h\to S_h$ the solution operator of this problem. This problem may also be written

$$u_{h,t} - \Delta_h u_h = 0, \quad \text{for } t\ge 0, \quad \text{with } u_h(0)=v_h,$$

where $\Delta_h : S_h \to S_h$ is now defined by

$$-(\Delta_h\psi,\chi)_h = (\nabla\psi,\nabla\chi), \quad \psi,\chi\in S_h. \tag{15.15}$$

With this notation $\bar E_h(t)$ is the semigroup on S_h generated by Δ_h.

Note that we may write the complex form of the discrete inner product $(\cdot,\cdot)_h$ defined in (15.5) as

$$(\psi,\chi)_h = \sum_{j=1}^{N_h}\omega_j\psi_j\bar\chi_j, \quad \text{where } \omega_j = \frac{1}{3}\sum_{P_j\in\bar\tau}\operatorname{area}(\tau), \quad \psi_j=\psi(P_j). \tag{15.16}$$

From this we may easily see that

$$-(\Delta_h \psi)_j = \omega_j^{-1} \sum_{i=1}^{N_h} \alpha_{ij} \psi_i, \quad \text{where } \alpha_{ij} = (\nabla \Phi_i, \nabla \Phi_j), \quad \text{for } j = 1, \dots, N_h.$$

We shall begin with a characterization of Delaunay triangulations.

Lemma 15.4 *The triangulation $\mathcal{T}_h$ is of Delaunay type if and only if*

$$(\nabla \Phi_i, \nabla \Phi_j) \le 0, \quad \textit{for all } P_i \ne P_j.$$

Proof. Let $e = P_i P_j$ be an edge of $\mathcal{T}_h$, let τ be one of the two triangles determined by e, and let α be the angle in τ opposite e. Then $\nabla \Phi_i\big|_\tau$ is in the direction of the normal to the side of τ opposite P_i and $|\nabla \Phi_i|\big|_\tau = 1/\delta_{i,\tau}$, where $\delta_{i,\tau}$ is the distance from P_i to the opposite side of τ. One sees at once that the angle between the two normals is $\pi - \alpha$, and hence

$$(\nabla \Phi_i, \nabla \Phi_j)_\tau = -\cos\alpha \, |\nabla \Phi_i|\big|_\tau \, |\nabla \Phi_j|\big|_\tau \, \text{area}(\tau) = -\cos\alpha \, \delta_{i,\tau}^{-1} \, \delta_{j,\tau}^{-1} \, \text{area}(\tau).$$

But we also have

$$\text{area}(\tau) = \ell_{i,\tau} \ell_{j,\tau} \, \sin\alpha = \ell_{i,\tau} \delta_{i,\tau}/2,$$

where $\ell_{i,\tau}$ is the length of the side opposite to P_i. Hence altogether

$$(\nabla \Phi_i, \nabla \Phi_j)_\tau = -\tfrac{1}{4} \cot\alpha.$$

We finally have

$$(\nabla \Phi_i, \nabla \Phi_j) = \sum_{j=1}^{2} (\nabla \Phi_i, \nabla \Phi_j)_{\tau_j} = -\tfrac{1}{4}(\cot\alpha_1 + \cot\alpha_2) = -\frac{\sin(\alpha_1 + \alpha_2)}{4 \sin\alpha_1 \sin\alpha_2},$$

from which the conclusion of the lemma immediately follows. □

We shall show the following discrete maximum-principle.

Theorem 15.5 *Assume that the triangulations $\mathcal{T}_h$ are of Delaunay type. Then*

$$\min(0, \min_{x \in \Omega} v_h(x)) \le (\bar{E}_h(t) v_h)(x) \le \max(0, \max_{x \in \Omega} v_h(x)), \quad \forall v_h \in S_h.$$

In particular, $\bar{E}_h(t)$ is stable with respect to the maximum-norm, and

$$\|\bar{E}_h(t) v_h\|_{L_\infty} \le \|v_h\|_{L_\infty}, \quad \textit{for } t \ge 0.$$

Proof. We write as above the system in matrix form

$$\bar{\mathcal{B}} \alpha'(t) + \mathcal{A} \alpha(t) = 0, \quad \text{for } t > 0, \quad \text{with } \alpha(0) = \gamma,$$

where $\alpha(t)$ and γ are the vectors whose components are the coefficients of $u_h(t) = \bar{E}_h(t)v_h$ and v_h with respect to the basis $\{\Phi_j\}_{j=1}^{N_h}$ of S_h, and where $\bar{\mathcal{B}} = ((\Phi_j, \Phi_l)_h)$ is diagonal and $\mathcal{A} = ((\nabla\Phi_j, \nabla\Phi_l))$ is the stiffness matrix. Clearly, the maxima and minima of $u_h(t)$ and v_h coincide with those of the components of $\alpha(t)$ and γ, respectively. Since, with $\widetilde{\mathcal{A}}$ and $\mathcal{G}(t)$ defined by the latter two equalities,

$$\alpha(t) = e^{-\bar{\mathcal{B}}^{-1}\mathcal{A}t}\gamma = e^{-\widetilde{\mathcal{A}}t}\gamma = \mathcal{G}(t)\gamma, \quad \text{for } t \geq 0,$$

it suffices for the first statement of the theorem to show that the matrix $\mathcal{G}(t) \geq 0$ (in the sense that its elements $g_{jl}(t)$ are nonnegative) and that for each j,

$$\sum_{l=1}^{N_h} g_{jl}(t) \leq 1. \tag{15.17}$$

For the purpose of showing $\mathcal{G}(t) \geq 0$ we observe that the off-diagonal elements of the stiffness matrix $\mathcal{A}$ are nonpositive by Lemma 15.4. Therefore we have use for the following simple matrix lemma.

Lemma 15.5 *Let $\mathcal{M} = (m_{jl})$ be a positive definite symmetric matrix with $m_{jl} \leq 0$ for $j \neq l$. Then $\mathcal{M}^{-1} \geq 0$.*

Proof. Let $\mu \geq \max_j m_{jj}$ be such that all eigenvalues of $\mathcal{K} = \mu\mathcal{I} - \mathcal{M}$ are positive. Then the largest eigenvalue of $\mathcal{K}$ and thus also its norm are smaller than μ. Hence

$$\mathcal{M}^{-1} = (\mu\mathcal{I} - \mathcal{K})^{-1} = \mu^{-1}(\mathcal{I} - \mu^{-1}\mathcal{K})^{-1} = \sum_{j=0}^{\infty} \mu^{-j-1}\mathcal{K}^j \geq 0,$$

since $\mathcal{K}$ has nonnegative elements. □

It follows from the lemma that $(\mathcal{I} + k\widetilde{\mathcal{A}})^{-1} \geq 0$ for $k > 0$. In fact, $\bar{\mathcal{B}} + k\mathcal{A}$ satisfies the assumptions of the lemma so that $(\bar{\mathcal{B}} + k\mathcal{A})^{-1} \geq 0$, and hence

$$(\mathcal{I} + k\widetilde{\mathcal{A}})^{-1} = (\bar{\mathcal{B}}^{-1}(\bar{\mathcal{B}} + k\mathcal{A}))^{-1} = (\bar{\mathcal{B}} + k\mathcal{A})^{-1}\bar{\mathcal{B}} \geq 0.$$

Since the powers of nonnegative matrices are nonnegative, we conclude

$$\mathcal{G}(t) = e^{-t\widetilde{\mathcal{A}}} = \lim_{n\to\infty} (\mathcal{I} + \frac{t}{n}\widetilde{\mathcal{A}})^{-n} \geq 0.$$

We now complete the proof by showing (15.17), that is, with $\underline{1}$ the N_h-vector with components 1, that (element-wise) $\mathcal{G}(t)\underline{1} \leq \underline{1}$. We shall show below that $\mathcal{A}\underline{1} \geq 0$. Assuming this for a moment we have $(\bar{\mathcal{B}} + k\mathcal{A})\underline{1} \geq \bar{\mathcal{B}}\underline{1}$. It follows that $(\bar{\mathcal{B}} + k\mathcal{A})^{-1}\bar{\mathcal{B}}\underline{1} = (I + k\widetilde{\mathcal{A}})^{-1}\underline{1} \leq \underline{1}$, and hence as above

$$\mathcal{G}(t)\underline{1} = e^{-t\widetilde{\mathcal{A}}}\underline{1} = \lim_{n\to\infty}(I + \frac{t}{n}\widetilde{\mathcal{A}})^{-n}\underline{1} \le \underline{1}.$$

For the purpose of showing that $\mathcal{A}\underline{1} \ge 0$, we extend the basis $\{\Phi_j\}_{j=1}^{N_h}$ with additional pyramid functions $\{\Phi_{N_h+l}\}_{l=1}^{M_h}$ corresponding to the boundary vertices. In fact, we only need to consider these defined on the polygonal domain Ω_h defined by $\mathcal{T}_h$, so no extension of $\mathcal{T}_h$ is needed. In the same way as before, we have for P_j an interior vertex and P_{N_h+l} a boundary vertex that $(\nabla\Phi_j, \nabla\Phi_{N_h+l}) \le 0$. Hence, since $\sum_{l=1}^{N_h+M_h} \Phi_l \equiv 1$ in Ω_h,

$$\sum_{l=1}^{N_h} a_{jl} = (\nabla\Phi_j, \nabla\sum_{l=1}^{N_h+M_h}\Phi_l) - \sum_{l=1}^{M_h}(\nabla\Phi_j, \nabla\Phi_{N_h+l}) \ge 0.$$

This shows $\mathcal{A}\underline{1} \ge 0$ and thus completes the proof of the maximum-principle. The second part of the theorem is an obvious consequence of the first. □

Maximum-principles are also valid under certain conditions for the homogeneous case ($f = 0$) of the fully discrete schemes (15.14) with $\kappa \in [0, 1]$. We show the following:

Theorem 15.6 *Assume that $\mathcal{T}_h$ is of Delaunay type, and that $(1-\kappa)k \le \delta_{\min}^2/3$, where $\delta_{\min} = \min_{j,\tau}\delta_{j,\tau}$. Then the solution of (15.14) with $f = 0$ satisfies, for $x \in \Omega$,*

$$\min(0, \min_{x\in\Omega} v_h(x)) \le U^n(x) \le \max(0, \max_{x\in\Omega} v_h(x)), \quad \textit{for } n \ge 0.$$

In particular,

$$\|U^n\|_{L_\infty} \le \|v_h\|_{L_\infty}.$$

Proof. We write the scheme (15.14) with $f = 0$ as above in matrix form,

$$\alpha^n = (\bar{\mathcal{B}} + \kappa k\mathcal{A})^{-1}(\bar{\mathcal{B}} - (1-\kappa)k\mathcal{A})\alpha^{n-1} = \bar{\mathcal{G}}_{k,\kappa}\alpha^{n-1}.$$

We need to show as before that $\bar{\mathcal{G}}_{k,\kappa} \ge 0$ and $\bar{\mathcal{G}}_{k,\kappa}\underline{1} \le \underline{1}$. For the backward Euler scheme, corresponding to $\kappa = 1$, our above proof of Theorem 15.5 shows the result. For more general $\kappa \in [0, 1]$ we still have $(\bar{\mathcal{B}}+\kappa k\mathcal{A})^{-1} \ge 0$ by Lemma 15.5. In order to guarantee $\bar{\mathcal{G}}_{k,\kappa} \ge 0$ we now demand $\bar{\mathcal{B}} - (1-\kappa)k\mathcal{A} \ge 0$. Since $a_{jl} \le 0$ for $j \ne l$ it suffices for this to require $\bar{b}_{jj} - (1-\kappa)ka_{jj} \ge 0$ or $(1-\kappa)k\|\nabla\Phi_j\|^2 \le \|\Phi_j\|_h^2$ for $j = 1, \dots, N_h$. But

$$\|\nabla\Phi_j\|^2 = \sum_{\tau\subset\operatorname{supp}\Phi_j} \delta_{j,\tau}^{-2} \text{ area } (\tau),$$

and, recalling that $D_j = \operatorname{supp}\Phi_j$,

$$\|\Phi_j\|_h^2 = \tfrac{1}{3}\sum_{\tau\in\operatorname{supp}\Phi_j} \text{area } (\tau) = \tfrac{1}{3} \text{ area } (D_j),$$

so that the condition is valid if $(1-\kappa)k \le \delta_{j,\tau}^2/3$, for all j, τ, which is satisfied under the assumptions of the theorem. Since $\mathcal{A}\underline{1} \ge 0$ as before we have $(\bar{\mathcal{B}} + k\kappa\mathcal{A})\underline{1} \ge (\bar{\mathcal{B}} - k(1-\kappa)\mathcal{A})\underline{1}$, and thus

$$\bar{\mathcal{G}}_{k,\kappa}\underline{1} = (\bar{\mathcal{B}} + k\kappa\mathcal{A})^{-1}(\bar{\mathcal{B}} - k(1-\kappa)\mathcal{A})\underline{1} \le \underline{1}.$$

This completes the proof of the theorem. □

Note that, except when $\kappa = 1$, a mesh-ratio condition of type $k \le Ch^2$ is required in this result.

We shall end this chapter by showing that the semigroup $\bar{E}_h(t)$ discussed above is, in fact, an analytic semigroup with respect to the maximum-norm. We shall then use this fact to conclude that it has a smoothing property and also to demonstrate a stability estimate for the fully discrete method (15.14). The analyticity of $\bar{E}_h(t)$ is a consequence of the following resolvent estimate, where we again assume that the family $\{\mathcal{T}_h\}$ is quasiuniform.

Theorem 15.7 *With Δ_h defined by* (15.15) *we have*

$$(15.18) \qquad \|R(z; -\Delta_h)\|_{L_\infty} \le C\ell_h^{1/2}|z|^{-1}, \quad \text{for } z \in \Sigma_{\delta_h},\ \delta_h = \tfrac{1}{2}\pi - c\ell_h^{-1/2}.$$

We begin by stating a resolvent estimate in the discrete L_p-norm which we define in analogy with (15.16) as

$$\|\chi\|_{L_p,h} = \Big(\sum_j \omega_j|\chi_j|^p\Big)^{1/p}, \quad \text{for } \chi \in S_h.$$

Theorem 15.8 *With Δ_h defined by* (15.15) *we have*

$$(15.19) \qquad \|R(z; -\Delta_h)\|_{L_p,h} \le \sqrt{p}\,|z|^{-1}, \quad \text{for } z \in \Sigma_{\delta_p},\quad \delta_p = \tfrac{1}{2}\pi - p^{-1/2}.$$

We now use this result to give the

Proof of Theorem 15.7. Setting $U = R(z; -\Delta_h)F$ we have, with j appropriate and $p < \infty$, since $\omega_j \ge ch^2$,

$$\|U\|_{L_\infty} = |U_j| \le \omega_j^{-1/p}\|U\|_{L_p,h} \le Ch^{-2/p}\sqrt{p}\,|z|^{-1}\|F\|_{L_p,h},$$

for $z \in \Sigma_{\delta_p}$, $\delta_p = \frac{1}{2}\pi - p^{-1/2}$. Choosing $p = \ell_h = \log(1/h)$ for small h now completes the proof. □

The basis of our L_p analysis is the following lemma where for an edge e of $\mathcal{T}_h$ defined by two neighbors P_i and P_j, $\partial_j U = U_{j_1} - U_{j_2}$.

Lemma 15.6 *For every edge e of $\mathcal{T}_h$ there is a real-valued constant γ_e such that*

$$(\nabla\psi, \nabla\chi) = \sum_j \gamma_e \partial_e\psi \cdot \overline{\partial_e\chi}, \quad \forall \psi, \chi \in S_h.$$

Proof. We note that $\sum_{i=1}^{N_h+M_h} \alpha_{ji} = 0$ since $\sum_{i=1}^{N_h+M_h} \Phi_i = 1$. It therefore suffices to remark that, noting that $\psi_j = \chi_j = 0$ for $N_h + 1 \le j \le N_h + M_h$,

$$(\nabla\psi, \nabla\chi) = \sum_{i,j=1}^{N_h+M_h} \alpha_{ij}\psi_i\bar{\chi}_j = \sum_{i\neq j} \alpha_{ij}(\psi_i - \psi_j)(\bar{\chi}_j - \bar{\chi}_i).$$

This shows the lemma with $\gamma_e = -a_{ij}$ for $e = P_iP_j$. □

We also need the following lemma:

Lemma 15.7 *Let z and w be two complex numbers and set*

$$H_p = (w - z)(\bar{w}|w|^{p-2} - \bar{z}|z|^{p-2}), \quad \text{where } p > 2.$$

Then

$$|\arg H_p| \le \arcsin(1 - 2/p).$$

Proof. Setting $d = w - z$ and $\varphi(t) = d\overline{(z + td)}\,|z + td|^{p-2}$ we may write

$$H_p = d\,\overline{(z + d)}\,|z + d|^{p-2} - d\,\bar{z}|z|^{p-2} = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)\,dt,$$

and it hence suffices to show $|\arg\varphi'(t)| \le \arcsin|1 - 2/p|$. For this we write $d^2\,\overline{(z + td)}^2 = re^{i\omega}$ we have

$$\begin{aligned}\varphi'(t) &= \frac{p}{2}|d|^2|z + td|^{p-2} + \frac{p-2}{2}d^2\,\overline{(z + td)}^2|z + td|^{p-4}\\ &= \tfrac{1}{2}|z + td|^{p-4}r^2(p + (p - 2)e^{2i\omega}).\end{aligned}$$

We now easily find

$$|\arg\varphi'(t)| = |\arg(p + (p - 2)e^{2i\omega})| \le \arcsin(1 - 2/p),$$

which completes the proof. □

Proof of Theorem 15.8. We first show, with $\theta_p = \frac{1}{2}\pi + \arcsin(1 - 2/p)$,

$$\|R(z; -\Delta_h)\|_{L_p,h} \le |z|^{-1}, \quad \text{for } z \in \Sigma_{\theta_p}. \tag{15.20}$$

Letting $U \in S_h$ be the solution of the discrete elliptic problem

$$zU + \Delta_h U = F, \tag{15.21}$$

we have $U = R(z; -\Delta_h)F$ so that the statement (15.20) will follow from

$$\|U\|_{L_p,h} \le |z|^{-1}\|F\|_{L_p,h}, \quad \text{for } z \in \Sigma_{\theta_p}. \tag{15.22}$$

We obtain from (15.21)

$$(15.23) \qquad -z(U,\chi)_h + (\nabla U, \nabla \chi) = -(F,\chi)_h, \quad \forall \chi \in S_h.$$

We choose $\chi = \widetilde{\chi} = I_h(U|U|^{p-2})$ and note that, by Lemma 15.6,

$$(\nabla U, \nabla \widetilde{\chi}) = \sum_e \gamma_e \partial_e U \partial_e(\bar{U}|U|^{p-2}) = \sum_e \gamma_e H_{p,e},$$

where, for $e = P_iP_j$,

$$H_{p,e} = (U_j - U_i)(\bar{U}_j|U_j|^{p-2} - \bar{U}_i|U_i|^{p-2}).$$

Note that each $H_{p,e}$ is of the form of H_p in Lemma 15.7, and this lemma therefore shows $|\arg(\nabla U, \nabla\widetilde{\chi})| \le \arcsin(1-2/p)$.

We may then write (15.23), with $\chi = \widetilde{\chi}$, as

$$(15.24) \qquad -z\|U\|^p_{L_p,h} + (\nabla U, \nabla \widetilde{\chi}) = -(F,\chi)_h.$$

and think of this as a relation of the form

$$(15.25) \qquad ae^{i\varphi} + be^{i\psi} = c, \quad \text{with } a, b > 0,\ \varphi,\ \psi \in \mathbb{R},$$

where $\varphi = \arg(-z)$ and where $|\arg\psi| \le \arcsin(1-2/p)$. By multiplication by $e^{-i\varphi}$ and taking real parts, this implies

$$(15.26) \qquad a \le |c|, \quad \text{if} \quad |\arg(-z)| \le \tfrac{1}{2}\pi - \arcsin(1-2/p),$$

since then $\cos(\psi - \varphi) \ge 0$. Hence

$$(15.27) \qquad |z|\,\|U\|^p_{L_p,h} \le \|F\|_{L_p,h}\|U\|^{p-1}_{L_p,h}, \quad \text{for } z \in \Sigma_{\theta_p},$$

from which (15.20) follows.

Noting that $\theta_p > \frac{1}{2}\pi$, we now want to derive a bound for the resolvent in a wider sector which extends to the right half-plane. For this we use (15.20) (with λ replaced by ζ) to obtain

$$\begin{aligned}\|R(z;-\Delta_h)\|_{L_p,h} &\le \|R(\zeta;-\Delta_h)\|_{L_p,h}/(1 - |z-\zeta|\|R(\zeta;-\Delta_h)\|_{L_p,h}) \\ &\le \frac{1}{|\zeta| - |z-\zeta|}, \quad \text{if } |\arg\zeta| = \theta_p, \quad |z-\zeta|/|\zeta| < 1.\end{aligned}$$

Letting $|\zeta| \to \infty$ we find $|\zeta| - |z-\zeta| \to |z|\cos(\theta_p - |\arg z|)$ and hence, with $M_p(\varphi) = 1/\cos(\theta_p - |\varphi|)$,

$$\|R(z;-\Delta_h)\|_{L_p,h} \le \frac{M_p(\arg z)}{|z|}, \quad \text{for } \theta_p - \tfrac{1}{2}\pi < |\arg z| \le \theta_p.$$

In particular, if we assume that $z \in \Sigma_{\pi/2 - \arcsin(1/\sqrt{p})}$, then

$$\cos(\theta_p - |\arg z|) \ge \cos(\arcsin(1/\sqrt{p}) + \arcsin(1-2/p)) = 1/\sqrt{p},$$

and hence

$$\|R(z;-\Delta_h)\|_{L_p,h} \le \sqrt{p}\,|z|^{-1}, \qquad \text{for } |\arg z| \ge \tfrac{1}{2}\pi - \arcsin(1/\sqrt{p}).$$

Since $\frac{1}{2}\pi - \arcsin(1/\sqrt{p}) \ge \frac{1}{2}\pi - 1/\sqrt{p}$, this shows (15.19) and thus completes the proof. □

In the same way as in Chapter 6, Theorem 15.7 can be translated into properties for the semigroup $\bar{E}_h(t) = e^{t\Delta_h}$. In particular, we have the following smoothing property in maximum-norm.

Theorem 15.9 *Assume that the $\mathcal{T}_h$ are of Delaunay type. Then we have*

$$\|\bar{E}_h'(t)\|_{L_\infty} \le C\ell_h t^{-1}, \quad \text{for } t > 0.$$

Proof. This follow at once from Theorem 15.7 and Lemma 6.6, with $M = M_h = C\ell_h^{1/2}$, $\delta = \delta_h = \frac{1}{2}\pi - c\ell_h^{-1/2}$, since

$$\cos\delta_h = \cos(\tfrac{1}{2}\pi - c\ell_h^{-1/2}) = \sin(c\ell_h^{-1/2}) \ge c\ell_h^{-1/2}, \quad c > 0. \qquad (15.28) \quad \square$$

Using the techniques of Chapter 9 one may also use the resolvent estimate of Theorem 15.7 to show stability of fully discrete methods. We illustrate with the homogeneous case of (15.14), which we now write

$$\bar{\partial}U^n - \kappa\Delta_h U^n + (1-\kappa)\Delta_h U^{n-1} = 0, \quad \text{for } n \ge 1, \quad \text{with } U^0 = v_h. \qquad (15.29)$$

The solution of this problem is then

$$U^n = E_{kh}^n v_h = r_\kappa(-\Delta_h)^n v_h, \quad \text{where } r_\kappa(z) = \frac{1-(1-\kappa)z}{1+\kappa z}. \qquad (15.30)$$

We first need a somewhat more precise result than that of Theorem 9.1 in the special case of the rational function in (15.30). Recalling from (6.48) that $\ell(t) = \max(1, \log(1/t))$ we have the following.

Lemma 15.8 *Let A be an operator in the Banach space $\mathcal{B}$ satisfying* (9.2) *and* (9.3)*, and let $r_\kappa(z)$ be the rational function in* (15.30) *with $\frac{1}{2} \le \kappa \le 1$. Then*

$$\|E_k^n\| \le CM\ell(\cos\delta), \quad \text{for } n \ge 0, \quad \text{with } E_k = r_\kappa(kA).$$

Proof. We follow the proof of Theorem 9.1. Consider first the case $\kappa > \frac{1}{2}$, so that $|r_\kappa(\infty)| < 1$. Choosing $\psi = \delta$ the estimates for the integrals over γ^R and $\gamma^{\varepsilon/n}$ are unchanged. For the integral over $\Gamma^R_{\varepsilon/n}$ we note that, as is readily proved,

$$|r_\kappa(z)| \le e^{-c\,\mathrm{Re}\, z} \le e^{-c\cos\delta\,|z|}, \quad \text{for } z \in \Sigma_\delta,\ |z| \le R,$$

and hence the bound in (9.11) is replaced by

$$\frac{M}{\pi}\int_{\varepsilon/n}^\infty e^{-cn\cos\delta\,\rho}\frac{d\rho}{\rho} \le CM\ell(\cos\delta).$$

For $\kappa = \frac{1}{2}$ we have $|r_\kappa(\infty)| = 1$, which case is handled correspondingly as in the proof of Theorem 9.1. $\square$

The following is now our stability result for (15.30).

Theorem 15.10 *Assume that the $\mathcal{T}_h$ are of Delaunay type. Then we have for the solution of the fully discrete scheme* (15.30)*, for* $\frac{1}{2} \le \kappa \le 1$,

$$\|U^n\|_{L_\infty} \le C\ell_h^{1/2}\ell(\ell_h)\|v_h\|_{L_\infty}, \quad \textit{for } n \ge 0.$$

Proof. Since $r_\kappa(z)$ is A-stable (cf. (9.5)) for $\frac{1}{2} \le \kappa \le 1$, this is an immediate consequence of Theorem 15.7 and Lemma 15.8, together with (15.28). □

The lumped mass method described here is a special case of a family of methods involving quadrature analyzed in Raviart [203]. The superconvergence result of Lemma 15.2 and the corresponding maximum-norm error estimate as well as the reduced smoothness estimates are from Chen and Thomée [49]. The maximum-principles of Theorems 15.5 and 15.6 are contained in Fujii [102], and applied in Ushijima [237], [238] to derive uniform convergence, which, except for the case of uniform triangulations, was only shown to be of first order in h. The resolvent estimate of Theorem 15.7 is from Crouzeix and Thomée [60], where also nonquasiuniform families of triangulations are considered. In Nie and Thomée [177] a lumped mass method with quadrature also in the other terms in the variational formulation was discussed for a nonlinear parabolic problem.