Chapter 4
Concepts for Discretized Problems

4.1 Convergence of Minimizers

We consider an abstract finite-dimensional minimization problem that seeks a minimizing function $u_h \in \mathscr{A}_h$ for a functional

$$I_h(u_h) = \int_\Omega W_h(\nabla u_h)\,\mathrm{d}x,$$

where the indices h in $\mathscr{A}_h$ and W_h refer to discretized versions of given counterparts in the infinite-dimensional variational problem for minimizing

$$I(u) = \int_\Omega W(\nabla u)\,\mathrm{d}x$$

in the set of functions $u \in \mathscr{A}$. We will often refer to the infinite-dimensional problem as the *continuous problem*, but this does not imply a continuity property of the functional or its integrand. The finite-dimensional problems will also be referred to as *discretized problems*. We recall that it is sufficient for the existence of discrete solutions to have coercivity and lower semicontinuity of I_h, while in the continuous situation, coercivity and the strictly stronger notion of weak lower semicontinuity of I are required. We discuss in this section the variational convergence of minimization problems and adopt concepts described in the textbook [5].

4.1.1 Failure of Convergence

A natural question to address is whether a family of discrete solutions $(u_h)_{h>0}$ converges to a minimizer $u \in \mathscr{A}$ for I with respect to some topology. Obviously, this requires the existence of a minimizer $u \in \mathscr{A}$ for I and convergence of the entire

S. Bartels, *Numerical Methods for Nonlinear Partial Differential Equations*,
Springer Series in Computational Mathematics 47,
DOI 10.1007/978-3-319-13797-1_4

sequence of approximations requires uniqueness of the continuous solution, or a certain selection principle contained in the discrete problems. Surprisingly, even if a solution exists for the continuous problem, if the discretization is conforming in the sense that $\mathscr{A}_h \subset \mathscr{A}$ and $W_h = W$, and if the family $(\mathscr{A}_h)_{h>0}$ is dense in $\mathscr{A}$, then convergence of discrete solutions may fail entirely.

Example 4.1 (*Lavrentiev phenomenon* [9]) Let $\mathscr{A}$ be the set of all functions $v \in W^{1,1}(0, 1)$ satisfying $v(0) = 0$ and $v(1) = 1$ and consider

$$I(u) = \int_0^1 \left(x - u^3\right)^2 |u'|^6 \,\mathrm{d}x.$$

For $h > 0$ let $\mathscr{T}_h$ be a triangulation of $(0, 1)$, and define $\mathscr{A}_h = \mathscr{A} \cap \mathscr{S}^1(\mathscr{T}_h)$. Then the function $u(x) = x^{1/3}$ is a minimizer for I in $\mathscr{A}$, but for every $h > 0$, we have

$$0 = \min_{u\in\mathscr{A}} I(u) < \min_{u\in\mathscr{A}\cap W^{1,\infty}(0,1)} I(u) \le \min_{u_h\in\mathscr{A}_h} I(u_h).$$

In particular, the discrete minimal energies cannot converge to the right value. The reason for this discrepancy is the incompatibility of the growth of the integrand of I and the exponent of the employed Sobolev space in the definition of $\mathscr{A}$.

The example shows that even the seemingly simple notion of convergence

$$\min_{u_h\in\mathscr{A}_h} I_h(u_h) \to \inf_{u\in\mathscr{A}} I(u)$$

for $h \to 0$ requires stronger arguments than just the density of the approximation spaces. Once convergence is understood, a natural question to investigate is whether a rate of convergence can be proved, i.e., whether there exists $\alpha > 0$ with

$$|\min_{u_h\in\mathscr{A}_h} I_h(u_h) - \inf_{u\in\mathscr{A}} I(u)| \le ch^\alpha.$$

Even if this is the case, it is not guaranteed that discrete solutions $u_h \in \mathscr{A}_h$ converge to a minimizer $u \in \mathscr{A}$ of I.

Example 4.2 (*Lack of weak lower semicontinuity*) Set $\mathscr{A} = W^{1,4}(0, 1)$ and let

$$I(u) = \int_0^1 \left(|u'|^2 - 1\right)^2 + u^4 \,\mathrm{d}x.$$

For $h > 0$ let $\mathscr{T}_h$ be a triangulation of $(0, 1)$ of maximal mesh-size h and define $\mathscr{A}_h = \mathscr{A} \cap \mathscr{S}^1(\mathscr{T}_h)$. Then $\inf_{u\in\mathscr{A}} I(u) = 0$ and

$$|\min_{u_h\in\mathscr{A}_h} I(u_h) - \inf_{u\in\mathscr{A}} I(u)| \le ch^4,$$

and any weakly convergent sequence of discrete minimizers $(u_h)_{h>0}$ satisfies $u_h \rightharpoonup 0$ in $W^{1,4}(\Omega)$ as $h \to 0$. Due to the nonconvexity of the integrand, we have that $u = 0$ is not a minimizer for I, i.e., $0 < 1 = I(0)$.

4.1.2 Γ-Convergence of Discretizations

The concept of Γ-convergence provides a concise framework to analyze convergence of a sequence of energy functionals and its minimizers. In an abstract form we consider a sequence of discrete minimization problems:

$$\text{Minimize } I_h(u_h) \text{ in the set of functions } u_h \in X_h.$$

Here, every space X_h is assumed to be a subspace of a Banach space X and I_h is allowed to attain the value $+\infty$, so that constraints contained in $\mathscr{A}_h \subset X_h$ can be incorporated in I_h. We formally extend the discrete problems to X by setting

$$I_h(u) = \begin{cases} I_h(u) & \text{if} \quad u \in X_h, \\ +\infty & \text{if} \quad u \notin X_h. \end{cases}$$

In the following, $h > 0$ stands for a sequence of positive real numbers that accumulate at zero.

Definition 4.1 Let X be a Banach space, $I : X \to \mathbb{R} \cup \{+\infty\}$, and let $(I_h)_{h>0}$ be a sequence of functionals $I_h : X \to \mathbb{R} \cup \{+\infty\}$. We say that *the sequence* $(I_h)_{h>0}$ *Γ-converges to I* as $h \to 0$, denoted by $I_h \to^{\Gamma} I$, with respect to a given topology ω on X if the following conditions hold:

(a) For every sequence $(u_h)_{h>0} \subset X$ with $u_h \to^{\omega} u$ for some $u \in X$, we have that $\liminf_{h\to 0} I_h(u_h) \geq I(u)$.
(b) For every $u \in X$ there exists a sequence $(u_h)_{h>0} \subset X$ with $u_h \to^{\omega} u$ and $I_h(u_h) \to I(u)$ as $h \to 0$.

Remark 4.1 The first condition is called liminf-inequality and implies that I is a lower bound for the sequence $(I_h)_{h>0}$ in the limit $h \to 0$. The second condition guarantees that the lower bound is attained, and the involved sequence is called a recovery sequence.

Unless otherwise stated, we consider the weak topology ω on X. For conforming discretizations, i.e., if $I_h(u_h) = I(u_h)$ for all $u_h \in X_h$, of well-posed minimization problems, a Γ-convergence result can be proved under moderate conditions.

Theorem 4.1 (Conforming discretizations) *Assume that $I_h(u_h) = I(u_h)$ for $u_h \in X_h$ and $h > 0$ and that the spaces $(X_h)_{h>0}$ are dense in X with respect to the strong topology of X. If I is weakly lower semicontinuous and strongly continuous, then we have $I_h \to^{\Gamma} I$ as $h \to 0$ with respect to weak convergence in X.*

Proof Let $(u_h)_{h>0} \subset X$ and $u \in X$ be such that $u_h \rightharpoonup u$ as $h \to 0$. To prove the liminf-inequality, we note that $I_h(u_h) \geq I(u_h)$ and thus the weak lower semicontinuity of I implies $\liminf_{h\to 0} I_h(u_h) \geq \liminf_{h\to 0} I(u_h) \geq I(u)$. To prove that $I(u)$ is attained for every $u \in X$, let $(u_h)_{h>0}$ be a sequence with $u_h \in X_h$ for every $h > 0$ and $u_h \to u$ in X. The strong continuity of I and $I_h(u_h) = I(u_h)$ imply that $I(u) = \lim_{h\to 0} I_h(u_h)$. □

The definition of Γ-convergence has remarkable consequences.

Proposition 4.1 (Γ-Convergence)

(i) *If $I_h \to^\Gamma I$ as $h \to 0$, then I is weakly lower semicontinuous on X.*
(ii) *If $I_h \to^\Gamma I$ as $h \to 0$ and for every $h > 0$ there exists $u_h \in X$ such that $I_h(u_h) \leq \inf_{v_h\in X} I_h(v_h) + \varepsilon_h$ with $\varepsilon_h \to 0$ as $h \to 0$ and $u_h \to^\omega u$ for some $u \in X$, then $I_h(u_h) \to I(u)$ and u is a minimizer for I.*
(iii) *If $I_h \to^\Gamma I$ and G is ω-continuous on X, then $I_h + G \to^\Gamma I + G$.*

Proof (i) Let $(u_j)_{j\in\mathbb{N}} \subset X$ be a sequence with $u_j \to^\omega u$ in X as $j \to \infty$. For every $j \in \mathbb{N}$ there exists a sequence $(u_j^h)_{h>0}$ such that $u_j^h \to^\omega u_j$ as $h \to 0$ and $I_h(u_j^h) \to I(u_j)$. For every $j \in \mathbb{N}$ we may thus choose $h_j > 0$, such that $|I(u_j) - I_{h_j}(u_j^{h_j})| \leq 1/j$ and $u_j^{h_j} \to^\omega u$ as $j \to \infty$. It follows that

$$I(u) \leq \liminf_{j\to\infty} I_{h_j}\big(u_j^{h_j}\big) = \liminf_{j\to\infty} I(u_j) - I(u_j) + I_{h_j}\big(u_j^{h_j}\big) = \liminf_{j\to\infty} I(u_j).$$

This proves the first statement.
(ii) If $u_h \to^\omega u$, then by condition (a) we have $I(u) \leq \liminf_{h\to 0} I_h(u_h)$. Moreover, due to (b) for every $v \in X$, there exists $(v_h)_{h>0} \subset X$ with $v_h \to^\omega v$ and $I_h(v_h) \to I(v)$ as $h \to 0$. Therefore, $I(u_h) \leq I(v_h) + \varepsilon_h$ and

$$I(u) \leq \liminf_{h\to 0} I_h(u_h) \leq \lim_{h\to 0} \big(I_h(v_h) + \varepsilon_h\big) = I(v),$$

i.e., u is a minimizer for I.
(iii) If G is ω-continuous, then $G(u_h) \to G(u)$ whenever $u_h \to^\omega u$ in X and the Γ-convergence of $I_h + G$ to $I + G$ follows directly from $I_h \to^\Gamma I$. □

4.1.3 *Examples of Γ-Convergent Discretizations*

We discuss some examples of Γ-convergence. As above, we always extend a functional I_h defined on a subspace $X_h \subset X$ by the value $+\infty$ to the whole space X.

Example 4.3 (*Poisson problem*) Let $X = H^1_{\mathrm{D}}(\Omega)$ and $X_h = \mathcal{S}^1_{\mathrm{D}}(\mathcal{T}_h)$ for a regular family of triangulations $(\mathcal{T}_h)_{h>0}$ of Ω. For $f \in L^2(\Omega)$ and $g \in L^2(\Gamma_{\mathrm{N}})$, let

$$I(u) = \frac{1}{2}\int_\Omega |\nabla u|^2 \,\mathrm{d}x - \int_\Omega f u \,\mathrm{d}x - \int_{\Gamma_{\mathrm{N}}} g u \,\mathrm{d}s$$

and let $I_h : H^1_D(\Omega) \to \mathbb{R} \cup \{+\infty\}$ coincide with I on $\mathscr{S}^1_D(\mathscr{T}_h)$. Since the Dirichlet energy is weakly lower semicontinuous and strongly continuous, the linear lower-order terms are weakly continuous on $H^1_D(\Omega)$, and since the finite element spaces are dense in $H^1_D(\Omega)$, we verify that $I_h \to^{\Gamma} I$ as $h \to 0$. Nonhomogeneous Dirichlet conditions can be included by considering the decomposition $u = \widetilde{u} + \widetilde{u}_D$ with $\widetilde{u} \in H^1_D(\Omega)$. For minimizers $u \in H^2(\Omega) \cap H^1_D(\Omega)$ of I and $u_h \in \mathscr{S}^1_D(\mathscr{T}_h)$ of I_h, we have

$$\big|I(u) - I_h(u_h)\big| \le ch.$$

A constant sequence of functionals can have a different Γ-limit.

Example 4.4 (*Relaxation*) For the sequence of functionals defined through $X = W^{1,4}(0,1)$,

$$I(u) = \int_0^1 \big(|u'|^2 - 1\big)^2 + u^4 \,\mathrm{d}x,$$

subspaces $X_h = \mathscr{S}^1(\mathscr{T}_h)$, and $I_h = I$ on X_h, we have that $I_h \to^{\Gamma} I^{**}$ in $W^{1,4}(0,1)$ with the convexified functional

$$I^{**}(u) = \int_0^1 \big(|u'|^2 - 1\big)_+^2 + u^4 \,\mathrm{d}x,$$

where $s_+ = \max\{s, 0\}$ for $s \in \mathbb{R}$. Since the integrand of I^{**} is convex, the functional is weakly lower semicontinuous. Using that $I_h(u_h) = I(u_h) \ge I^{**}(u_h)$ for all $h > 0$, we deduce that $\liminf_{h\to 0} I_h(u_h) \ge I^{**}(u)$ whenever $u_h \rightharpoonup u$ in $W^{1,4}(0,1)$. To prove that the lower bound is attained, we first consider the case that $u \in W^{1,4}(\Omega)$ is piecewise affine, i.e., $u = u_H \in \mathscr{S}^1(\mathscr{T}_H)$ for some $H > 0$. For $0 < h < H$ we then construct a function u_h that nearly coincides with u_H on elements $T_H \in \mathscr{T}_H$ for which $|u'_H|_{T_H}| \ge 1$. For elements with $|u'_H|_{T_H}| \le 1$ we use gradients $u'_h \in \{\pm 1\}$ on T_H in such a way that u_h and u_H nearly coincide at the endpoints of T_H and differ by at most h in the interior. Then $I(u_h) \approx I^{**}(u_H)$ and $I(u_h) \to I^{**}(u_H)$ as $h \to 0$. The construction is depicted in Fig. 4.1. The assertion for general $u \in W^{1,4}(\Omega)$ follows from an approximation result and the strong continuity of I.

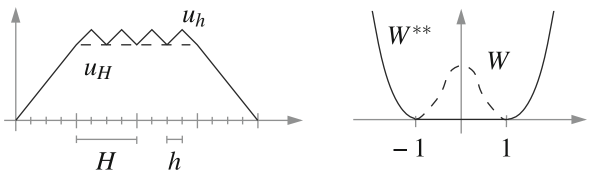

Fig. 4.1 Construction of an oscillating function u_h (*solid line*) with $|u'_h| \ge 1$ that approximates u_H (*dashed line*) such that $I(u_h) \approx I^{**}(u_H)$ (*left*) in Example 4.4; the integrand W^{**} (*solid line*) of I^{**} is the convex hull of the integrand W (*dashed line*) of I (*right*)

A typical application of conforming discretizations of well-posed minimization problems occurs in simulating hyperelastic materials.

Example 4.5 (*Hyperelasticity*) Let $\mathscr{A} = \{y \in W^{1,p}(\Omega;\mathbb{R}^d) : y|_{\Gamma_\mathrm{D}} = \widetilde{y}_\mathrm{D}|_{\Gamma_\mathrm{D}}\}$ for $1 \le p < \infty$ and $\widetilde{y}_\mathrm{D} \in W^{1,p}(\Omega;\mathbb{R}^d)$. Assume that $W : \mathbb{R}^{d\times d} \to \mathbb{R}$ is continuous and quasiconvex with

$$-c_1 + c_2|F|^p \le W(F) \le c_1 + c_2|F|^p.$$

Then for $f \in L^{p'}(\Omega;\mathbb{R}^d)$ and $g \in L^{p'}(\Gamma_\mathrm{N};\mathbb{R}^d)$, the functional

$$I(y) = \int_\Omega W(\nabla y)\,\mathrm{d}x - \int_\Omega f\cdot y\,\mathrm{d}x - \int_{\Gamma_\mathrm{N}} g\cdot y\,\mathrm{d}s$$

is weakly lower semicontinuous and coercive on $W^{1,p}(\Omega;\mathbb{R}^d)$. Moreover, if the sequence $(y_j)_{j\in\mathbb{N}} \subset W^{1,p}(\Omega;\mathbb{R}^d)$ converges strongly to $y \in W^{1,p}(\Omega;\mathbb{R}^d)$ then we have $\nabla y_{j_k}(x) \to \nabla y(x)$ for almost every $x \in \Omega$ for a subsequence $(y_{j_k})_{k\in\mathbb{N}}$, and the generalized dominated convergence theorem implies

$$\int_\Omega W(\nabla y_{j_k})\,\mathrm{d}x \to \int_\Omega W(\nabla y)\,\mathrm{d}x,$$

i.e., up to subsequences I is strongly continuous and this is sufficient to establish Γ-convergence. For piecewise affine boundary data y_D, we have that $\mathscr{A}_h = \mathscr{A} \cap \mathscr{S}^1(\mathscr{T}_h)^d$ is nonempty and the density of finite element spaces implies $I_h \to^\Gamma I$ for conforming discretizations. More generally, it suffices to consider convergent approximations $\widetilde{y}_{\mathrm{D},h}$ of $\widetilde{y}_\mathrm{D}$.

The abstract convergence theory allows us to include nonlinear constraints.

Example 4.6 (*Harmonic maps*) Assume that $u_\mathrm{D} \in C(\Gamma_\mathrm{D};\mathbb{R}^m)$ is such that

$$\mathscr{A} = \{u \in H^1(\Omega;\mathbb{R}^m) : u|_{\Gamma_\mathrm{D}} = u_\mathrm{D},\ |u(x)| = 1 \text{ f.a.e. } x \in \Omega\}$$

is nonempty and for a triangulation $\mathscr{T}_h$ of Ω with nodes $\mathscr{N}_h$, set

$$\mathscr{A}_h = \{u_h \in \mathscr{S}^1(\mathscr{T}_h)^m : u(z) = u_\mathrm{D}(z) \text{ f.a. } z \in \mathscr{N}_h\cap\Gamma_\mathrm{D},\ |u_h(z)| = 1 \text{ f.a. } z \in \mathscr{N}_h\},$$

i.e., $\mathscr{A}_h \not\subset \mathscr{A}$. We then consider the minimization of the Dirichlet energy I on $\mathscr{A}_h$ and $\mathscr{A}$, respectively, which defines minimization problems with functionals I_h and I on $H^1(\Omega;\mathbb{R}^m)$, respectively. To show that $I_h \to^\Gamma I$ in $H^1(\Omega;\mathbb{R}^m)$ we note that the liminf-inequality follows from the weak lower semicontinuity of I, together with the fact that if $u_h \rightharpoonup u$ in $W^{1,2}(\Omega;\mathbb{R}^m)$ with $u_h \in \mathscr{A}_h$ for every $h > 0$, then $u \in \mathscr{A}$. The latter implication follows from a nodal interpolation result, together with elementwise inverse estimates, i.e.,

$$\||u_h|^2 - 1\| = \||u_h|^2 - \mathscr{I}_h|u_h|^2\| \le ch\|u_h\|\|\nabla u_h\|.$$

Therefore, $|u_{h'}(x)| \to 1$ for almost every $x \in \Omega$ and a subsequence $h' > 0$ so that $|u(x)| = 1$ for almost every $x \in \Omega$. We assume that u_{D} is sufficiently regular, so that a similar argument shows $u|_{\Gamma_{\mathrm{D}}} = u_{\mathrm{D}}$. To prove the attainment of I, we note that due to the density of smooth unit-length vector fields in $\mathscr{A}$, we may assume $u \in \mathscr{A} \cap H^2(\Omega; \mathbb{R}^m)$ and define $u_h = \mathscr{I}_h u \in \mathscr{A}_h$. Then $u_h \to u$ in $H^1(\Omega; \mathbb{R}^m)$ and $I_h(u_h) \to I(u)$ as $h \to 0$.

Remark 4.2 In general, smooth constrained vector fields are not dense in sets of weakly differentiable constrained vector fields, cf., e.g., [18].

For practical purposes it is often desirable to modify a given functional.

Example 4.7 (*Total variation minimization*) For $X = W^{1,1}(\Omega)$ we consider

$$I(u) = \int_\Omega |\nabla u| \, \mathrm{d}x;$$

and given a family of triangulations $(\mathscr{T}_h)_{h>0}$ of Ω and $u_h \in \mathscr{S}^1(\mathscr{T}_h)$, we define for $\beta > 0$ the regularized functionals

$$I_h(u_h) = \int_\Omega (h^\beta + |\nabla u_h|^2)^{1/2} \, \mathrm{d}x.$$

If $u_h \rightharpoonup u$ in $W^{1,1}(\Omega)$, then the liminf-inequality follows from the weak lower semicontinuity of I on $W^{1,1}(\Omega)$ and the fact that $I_h(u_h) \geq I(u_h)$ for every $h > 0$. To verify that $I(u)$ is attained for every $u \in W^{1,1}(\Omega)$ in the limit $h \to 0$, we note that the density of finite element spaces in $W^{1,1}(\Omega)$ allows us to consider a sequence $(u_h)_{h>0} \subset W^{1,1}(\Omega)$ with $u_h \in \mathscr{S}^1(\mathscr{T}_h)$ for every $h > 0$ and $u_h \to u \in W^{1,1}(\Omega)$ as $h \to 0$. The estimate $(a^2 + b^2)^{1/2} \leq |a| + |b|$ implies that

$$(h^\beta + |\nabla u_h|^2)^{1/2} - |\nabla u| \leq h^{\beta/2} + |\nabla u_h| - |\nabla u|,$$

and for a subsequence we have $((h')^\alpha + |\nabla u_{h'}|^2)^{1/2} \to |\nabla u|$ almost everywhere in Ω. The generalized dominated convergence theorem implies that $I_{h'}(u_{h'}) \to I(u)$ as $h' \to 0$. With Proposition 4.1, this also implies the Γ-convergence of discretizations of

$$I(u) = \int_\Omega |\nabla u| \, \mathrm{d}x + \frac{\alpha}{2} \|u - g\|^2$$

for $g \in L^2(\Omega)$. Due to the lack of reflexivity of $W^{1,1}(\Omega)$ this is not sufficient to deduce the existence of minimizers for I, i.e., we cannot deduce the existence of weak limits of (subsequences) of a bounded sequence. For this, the larger space $BV(\Omega) \cap L^2(\Omega)$ has to be considered. A corresponding Γ-convergence result follows analogously with the density of $W^{1,1}(\Omega)$ in $BV(\Omega)$ with respect to an appropriate notion of convergence.

4.1.4 Error Control for Strongly Convex Problems

For Banach spaces X and Y, a bounded linear operator $\Lambda : X \to Y$, and convex, lower-semicontinuous, proper functionals $F : X \to \mathbb{R} \cup \{+\infty\}$ and $G : Y \to \mathbb{R} \cup \{+\infty\}$, we consider the problem of finding $u \in X$ with

$$I(u) = \inf_{v \in X} I(v), \quad I(v) = F(v) + G(\Lambda v).$$

The *Fenchel conjugates* $F^* : X' \to \mathbb{R} \cup \{+\infty\}$ and $G^* : Y' \to \mathbb{R} \cup \{+\infty\}$ are the convex, lower-semicontinuous, proper functionals defined by

$$F^*(w) = \sup_{v \in X} \langle w, v \rangle - F(v), \quad G^*(q) = \sup_{p \in Y} \langle q, p \rangle - G(p)$$

for $w \in X'$ and $q \in Y'$, respectively. We assume that Y is reflexive, so that $G = G^{**}$. Then, the property of the formal adjoint operator $\Lambda' : Y' \to X'$, that $\langle \Lambda v, q \rangle = \langle v, \Lambda' q \rangle$, and the general relation $\inf_v \sup_q H(v, q) \geq \sup_q \inf_v H(v, q)$ for an arbitrary function $H : X \times Y' \to \mathbb{R} \cup \{+\infty\}$ yield

$$\begin{aligned}
\inf_v I(v) = \inf_v F(v) + G^{**}(\Lambda v) &= \inf_v \sup_q F(v) + \langle v, \Lambda' q \rangle - G^*(q) \\
&\geq \sup_q \inf_v F(v) + \langle v, \Lambda' q \rangle - G^*(q) = \sup_q \inf_v F(v) - \langle v, -\Lambda' q \rangle - G^*(q) \\
&= \sup_q \Big(- \sup_v \langle v, -\Lambda' q \rangle - F(v) - G^*(q) \Big) = \sup_q -F^*(-\Lambda' q) - G^*(q).
\end{aligned}$$

This motivates considering the dual problem which consists in finding $p \in Y'$ with

$$D(p) = \sup_{q \in Y'} D(q), \quad D(q) = -F^*(-\Lambda' q) - G^*(q).$$

We assume that F or G is *strongly convex*, so that there exist $\alpha_F, \alpha_G \geq 0$ with $\max\{\alpha_F, \alpha_G\} > 0$, so that for all $q_1, q_2 \in Y$ and $v_1, v_2 \in X$, we have

$$\begin{aligned}
G\big((q_1 + q_2)/2\big) + \alpha_G \|q_2 - q_1\|_Y^2 &\leq \frac{1}{2}\big(G(q_1) + G(q_2)\big), \\
F\big((v_1 + v_2)/2\big) + \alpha_F \|v_2 - v_1\|_X^2 &\leq \frac{1}{2}\big(F(v_1) + F(v_2)\big).
\end{aligned}$$

By convexity, the estimates hold with $\alpha_G = \alpha_F = 0$. The primal and dual optimization problems are related by the weak *complementarity principle*

$$I(u) = \inf_{v \in X} I(v) \geq \sup_{q \in Y^*} D(q) = D(p).$$

We say that *strong duality* applies if equality holds. Our final ingredient for the error estimate is a characterization of the optimality of the solution of the primal problem.

For some $\alpha_I \geq 0$ and all $w \in \partial I(u)$, we have that

$$\langle w, v-u\rangle + \alpha_I \|v-u\|_X^2 \leq I(v) - I(u)$$

and u is optimal if and only if $0 \in \partial I(u)$. We assume in the following that $\alpha_F > 0$ or $\alpha_I > 0$, so that I has a unique minimizer $u \in X$.

Theorem 4.2 (Error control [16]) *Assume that* $\max\{\alpha_F, \alpha_G, \alpha_I\} > 0$ *and let* $u \in X$ *be the unique minimizer for* I.

(i) *For a minimizer* $u_h \in X_h$ *for* I *restricted to a subspace* $X_h \subset X$*, we have the a priori error estimate*

$$\alpha_G \|\Lambda(u-u_h)\|_Q^2 + (\alpha_F + \alpha_I/4)\|u-u_h\|_X^2 \leq \inf_{w_h \in X_h} \frac{1}{2}\big(I(w_h) - I(u)\big).$$

(ii) *For an arbitrary approximation* $\widetilde{u}_h \in X$ *of* u*, we have the a posteriori error estimate*

$$\alpha_G \|\Lambda(u-\widetilde{u}_h)\|_Q^2 + (\alpha_F + \alpha_I/4)\|u-\widetilde{u}_h\|_X^2 \leq \inf_{q \in Y'} \frac{1}{2}\big(I(\widetilde{u}_h) - D(q)\big).$$

Proof The convexity estimates imply that

$$\alpha_G \|\Lambda(u-v)\|_Q^2 + \alpha_F \|u-v\|_X^2 \leq \frac{1}{2}\big(I(v) + I(u)\big) - I\big((v+u)/2\big).$$

The optimality of u shows that we have

$$I(u) + \alpha_I \|u-(u+v)/2\|_X^2 \leq I\big((u+v)/2\big).$$

It follows that

$$\alpha_G \|\Lambda(u-v)\|_Q^2 + \alpha_F \|u-v\|_X^2 \leq \frac{1}{2}\big(I(v) - I(u)\big) - \alpha_I \|((u-v)/2\|_X^2.$$

If $u_h \in X_h$ is minimal in X_h, then the identity $I(u_h) = \inf_{w_h \in X_h} I(w_h)$ implies the a priori estimate. The weak complementarity principle $I(u) \geq D(q)$ yields the a posteriori estimate. $\square$

Remarks 4.3 (i) If *strong duality* holds, i.e., if $I(u) = D(p)$, then the estimate of the theorem is sharp in the sense that the right-hand side vanishes if $v = u$ and q solves the dual problem.
(ii) Sufficient conditions for strong duality are provided by von Neumann's minimax theorem, e.g., that F and G^* are convex, lower semicontinuous, and coercive.

Example 4.8 For the Poisson problem $-\Delta u = f$ in Ω, $u|_{\partial\Omega} = 0$, we have $X = H_0^1(\Omega)$, $Y = L^2(\Omega;\mathbb{R}^d)$, $\Lambda = \nabla$, $G(\Lambda v) = (1/2)\int_\Omega |\nabla v|^2 \,\mathrm{d}x$, and

$F(v) = -\int_\Omega f v \,\mathrm{d}x$. It follows that $F^*(w) = I_{\{-f\}}(w)$, $G^*(q) = (1/2)\int_\Omega |q|^2 \,\mathrm{d}x$,

$$\Lambda' = -\operatorname{div} : L^2(\Omega; \mathbb{R}^d) \to H_0^1(\Omega)^*.$$

We thus have

$$\frac{1}{2}\big((q_1+q_2)/2\big)^2 - \frac{1}{4}(q_1^2+q_2^2) = \frac{1}{8}(q_1^2+2q_1q_2+q_2^2-2q_1^2-2q_2^2) = -\frac{1}{8}(q_1-q_2)^2,$$

so that $\alpha_G = 1/8$ and

$$\frac{1}{2}q_1^2 - \frac{1}{2}q_2^2 - q_1(q_1-q_2) = -\frac{1}{2}q_1^2 - \frac{1}{2}q_2^2 + q_1q_2 = -\frac{1}{2}(q_1-q_2)^2,$$

i.e., $\alpha_I = 1/2$. Moreover, we have $\alpha_F = 0$.

(i) Incorporating the definition of the exact weak solution, the abstract a priori estimate of Theorem 4.2 provides the bound

$$\begin{aligned}\frac{1}{2}\|\nabla(u-u_h)\|^2 &\le \frac{1}{2}\int_\Omega |\nabla w_h|^2 - \int_\Omega f w_h \,\mathrm{d}x - \frac{1}{2}\int_\Omega |\nabla u|^2 + \int_\Omega f u \,\mathrm{d}x \\ &= \frac{1}{2}\|\nabla(u-w_h)\|^2 + \int_\Omega \nabla u \cdot \nabla(u-w_h)\,\mathrm{d}x + \int_\Omega f(u-w_h)\,\mathrm{d}x \\ &= \frac{1}{2}\|\nabla(u-w_h)\|^2,\end{aligned}$$

which implies the best-approximation property

$$\|\nabla(u-u_h)\| \le \inf_{w_h\in X_h} \|\nabla(u-w_h)\|.$$

(ii) Letting $\eta^2(v,q)$ denote the right-hand side of the a posteriori error estimate of Theorem 4.2, we have

$$\begin{aligned}2\eta^2(v,q) &= -\int_\Omega f v \,\mathrm{d}x + I_{\{-f\}}(\operatorname{div} q) + \frac{1}{2}\int_\Omega |\nabla v|^2 \,\mathrm{d}x + \frac{1}{2}\int_\Omega |q|^2 \,\mathrm{d}x \\ &= \int_\Omega (\operatorname{div} q)v \,\mathrm{d}x + \frac{1}{2}\|\nabla v\|^2 + \frac{1}{2}\|q\|^2 = \frac{1}{2}\|\nabla v - q\|^2,\end{aligned}$$

provided that $-\operatorname{div} q = f$. The theorem thus implies

$$\|\nabla(u-v)\| \le \inf_{-\operatorname{div} q = f} \|\nabla v - q\|.$$

4.2 Approximation of Equilibrium Points

The Euler–Lagrange equations related to a minimization problem typically seek a function $u \in X$ such that

$$F(u)[v] = \ell(v)$$

for all $v \in X$ with a possibly nonlinear operator $F : X \to X'$ and a linear functional $\ell \in X'$. Various other mathematical problems that may not be related to a minimization problem can also be formulated in this abstract form. A natural discretization employs subspaces $X_h \subset X$ and seeks $u_h \in X_h$ with

$$F_h(u_h)[v_h] = \ell_h(v_h)$$

for all $v_h \in X_h$. Here, $F_h : X_h \to X_h'$ and $\ell_h \in X_h'$ are approximations of F and ℓ that result from a discretization, e.g., via numerical integration. The important question to address is whether numerical solutions $(u_h)_{h>0}$ for a sequence of finite-dimensional subspaces X_h converge in an appropriate sense to a solution of the infinite-dimensional problem. We assume that the finite-dimensional space X_h is equipped with the norm of X. The corresponding dual spaces X_h' and X' are related by the inclusion $X'|_{X_h} \subset X_h'$. Topics related to the contents of this section can be found in the textbooks [3, 11].

4.2.1 Failure of Convergence

The following examples show that unjustified regularity assumptions can lead to the failure of convergence to the correct object. The following examples are taken from [6].

Example 4.9 (*Maxwell's equations*) For $\Omega \subset \mathbb{R}^2$ set $X = H_0(\text{curl}; \Omega) \cap H(\text{div}; \Omega)$, where

$$H_0(\text{curl}; \Omega) = \{v \in L^2(\Omega; \mathbb{R}^2) : \text{curl } v \in L^2(\Omega; \mathbb{R}^2),\ v \cdot t = 0 \text{ on } \partial\Omega\}$$

with $\text{curl } v = \partial_1 v_2 - \partial_2 v_1$ for $v = (v_1, v_2)$ and $t : \partial\Omega \to \mathbb{R}^2$ a unit tangent. For $f \in L^2(\Omega; \mathbb{R}^2)$, consider the problem of finding $u \in X$ such that

$$(\text{curl } u, \text{curl } v) + (\text{div } u, \text{div } v) = (f, v)$$

for all $v \in X$. The existence and uniqueness of a solution follows from the Lax–Milgram lemma. A discretization of this problem is obtained by choosing $X_h = \mathscr{S}^1(\mathscr{T}_h)^2 \cap X$ and computing $u_h \in X_h$ such that

$$(\text{curl } u_h, \text{curl } v_h) + (\text{div } u_h, \text{div } v_h) = (f, v_h)$$

for all $v_h \in X_h$. This defines a convergent numerical scheme if Ω is convex. If Ω is nonconvex, then $H^1(\Omega; \mathbb{R}^2) \cap X$ is a closed proper subspace of X, cf. [8] for details, and convergence $u_h \to u$ as $h \to 0$ fails in general.

A similar effect occurs for higher-order problems.

Example 4.10 (*Biharmonic equation*) The biharmonic equation

$$\Delta^2 u = f \text{ in } \Omega, \quad u = \Delta u = 0 \quad \text{on } \partial\Omega$$

formally corresponds to the weak formulation that seeks $u \in H^2(\Omega) \cap H^1_0(\Omega)$ with

$$\int_\Omega D^2 u : D^2 v \,\mathrm{d}x = \int_\Omega f v \,\mathrm{d}x$$

for all $v \in H^2(\Omega) \cap H^1_0(\Omega)$ We denote the unique weak solution of the variational formulation by $u = (\Delta^2)^{-1} f$. A natural discretization of the problem is based on an operator splitting which is obtained by introducing $z = -\Delta u$ and solving the Poisson problems

$$\begin{aligned} -\Delta z &= f \text{ in } \Omega, \quad z = 0 \text{ on } \partial\Omega, \\ -\Delta u &= z \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega. \end{aligned}$$

We have $z = (-\Delta)^{-1} f$ and $u = (-\Delta)^{-1} z = (-\Delta)^{-2} f$. Unless Ω is convex so that $\Delta u \in H^1_0(\Omega)$ we do not have $(\Delta^2)^{-1} f = (-\Delta)^{-2} f$, and convergence of related numerical methods will fail in general.

Failure of convergence may also be related to the lack of uniqueness of a solution as in the case of degenerately monotone problems.

Example 4.11 (*Degenerate monotonicity*) For $\sigma(F) = DW^{**}(F)$ for $F \in \mathbb{R}^d$ and $W^{**}(F) = (|F|^2 - 1)^2_+$, there are infinitely many functions $u \in W^{1,4}_0(\Omega)$ satisfying $F(u)[v] = \int_\Omega \sigma(\nabla u) \cdot \nabla v \,\mathrm{d}x = 0$ for all $v \in W^{1,4}_0(\Omega)$.

4.2.2 Abstract Error Estimates

We sketch below the classical concept that consistency and stability imply the convergence of numerical approximations, provided that appropriate regularity results are available. Dual to this is an approach that leads to computable upper bounds for the approximation error and which avoids regularity assumptions entirely.

Theorem 4.3 (Abstract a priori error estimate) *Let $u \in X$ satisfy $F(u) = \ell$ and assume that for an interpolant $i_h u \in X_h$ and a consistency functional $\mathscr{C}_h(u) \in X'_h$, we have*

$$F_h(i_h u)[v_h] - \ell_h(v_h) = \mathscr{C}_h(u; v_h)$$

for all $v_h \in X_h$. Assume that we have discrete stability in the sense that for all $z_h \in X_h$ and $b_h \in X'_h$, the implication

$$\forall v_h \in X_h \;\; F_h(z_h)[v_h] = b_h(v_h) \implies \|z_h\|_X \le c_{S,h} \|b_h\|_{X'_h}$$

holds. Then, if $F_h : X_h \to X_h'$ is linear, there exists a unique solution $u_h \in X_h$ with

$$\|u_h - i_h u\|_X \le c_{S,h} \|\mathscr{C}_h(u)\|_{X_h'}.$$

Proof Discrete stability implies that $F_h : X_h \to X_h'$ is a bijection and hence there exists a unique $u_h \in X_h$ with $F_h(u_h) = 0$. Since $F_h(i_h u - u_h) = F_h(i_h u) - F_h(u_h) = F_h(i_h u) - \ell_h = \mathscr{C}_h(u)$ we deduce the estimate. □

Remark 4.4 We say that a discretization is consistent of order $\beta \ge 0$, given the regularity $u \in Z \subset X$ if $\|\mathscr{C}_h(u)\|_{X_h'} \le ch^\beta$. This implies convergence of approximations with rate β.

A similar abstract concept leads to a posteriori error estimates for many linear problems.

Theorem 4.4 (Abstract a posteriori error estimate) *Let $u_h \in X_h$ and define the residual $\mathscr{R}_h(u_h) \in X'$ through*

$$\mathscr{R}_h(u_h; v) = F(u_h)[v] - \ell(v)$$

for all $v \in X$. Assume that we have the continuous stability result that for all $z \in X$ and $b \in X'$, the implication

$$\forall v \in X \ \ F(z)[v] = b(v) \implies \|z\|_X \le c_S \|b\|_{X'}$$

holds. If $u \in X$ satisfies $F(u) = \ell$ and if F is linear, then u is unique with

$$\|u - u_h\|_X \le c_S \|\mathscr{R}_h(u_h)\|_{X'}.$$

Proof The difference $u - u_h$ satisfies $F(u - u_h)[v] = \mathscr{R}_h(u_h; v)$ for all $v \in X$, and the stability result implies the error estimate and the uniqueness property. □

Example 4.12 (*Poisson problem*) Let $u \in H^1_\mathrm{D}(\Omega)$ be the weak solution of $-\Delta u = f$ in Ω, $u|_{\Gamma_\mathrm{D}} = 0$, and $\partial_\nu u|_{\Gamma_\mathrm{N}} = g$, i.e., we have $F(u) = \ell$ with

$$F(u)[v] = \int_\Omega \nabla u \cdot \nabla v \,\mathrm{d}x, \quad \ell(v) = \int_\Omega f v \,\mathrm{d}x + \int_{\Gamma_\mathrm{N}} g v \,\mathrm{d}s.$$

The lowest-order finite element method seeks $u_h \in \mathscr{S}^1_\mathrm{D}(\mathscr{T}_h)$ with $F(u_h)[v_h] = \ell(v_h)$ for all $v_h \in \mathscr{S}^1_\mathrm{D}(\mathscr{T}_h)$.

(i) Inserting an interpolant $i_h u \in \mathscr{S}^1_\mathrm{D}(\mathscr{T}_h)$ in the discrete formulation leads to

$$\mathscr{C}_h(u; v_h) = F(i_h u)[v_h] - \ell(v_h) = \int_\Omega \nabla[i_h u - u] \cdot \nabla v_h \,\mathrm{d}x$$

for all $v_h \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)$. We have $\|\mathscr{C}_h(u)\|_{\mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)'} \le ch\|D^2u\|$ if $u \in H^2(\Omega)\cap H^1_{\mathrm{D}}(\Omega)$ and $i_h u = \mathscr{I}_h u$ is the nodal interpolant of u. If $z_h \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)$ and $b_h \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)'$ are such that

$$\int_\Omega \nabla z_h \cdot \nabla v_h \,\mathrm{d}x = b_h(v_h)$$

for all $v_h \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)$, then the choice of $v_h = z_h$ shows the discrete stability estimate $\|\nabla z_h\| \le \|b_h\|_{\mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)'}$. Therefore, Theorem 4.3 implies the error estimate

$$\|\nabla(u_h - \mathscr{I}_h u)\|_{L^2(\Omega)} \le ch\|D^2u\|_{L^2(\Omega)}.$$

(ii) Let $u_h \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)$ and define

$$\mathscr{R}_h(u_h; v) = F(u_h)[v] - \ell(v) = \int_\Omega \nabla u_h \cdot \nabla v \,\mathrm{d}x - \int_\Omega f v \,\mathrm{d}x - \int_{\Gamma_{\mathrm{N}}} g v \,\mathrm{d}s$$

for all $v \in H^1_{\mathrm{D}}(\Omega)$. Noting the stability estimate $\|\nabla z\| \le \|b\|_{X'}$ for $z \in H^1_{\mathrm{D}}(\Omega)$ and $b \in H^1_{\mathrm{D}}(\Omega)'$ with

$$\int_\Omega \nabla z \cdot \nabla v \,\mathrm{d}x = b(v)$$

for all $v \in H^1_{\mathrm{D}}(\Omega)$, Theorem 4.4 implies the error estimate

$$\|\nabla(u - u_h)\|_{L^2(\Omega)} \le \|\mathscr{R}_h(u_h)\|_{X'}.$$

If u_h satisfies $F(u_h)[v_h] = 0$ for all $v_h \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)$, we have the Galerkin orthogonality $F(u - u_h)[v_h] = 0$ for all $v_h \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)$ and $\|\mathscr{R}_h(u_h)\|_{X'} \le c\eta(u_h)$ with a computable quantity $\eta(u_h)$, cf. Theorem 3.6.

The concepts can be generalized to the class of strongly monotone operators.

Definition 4.2 The operator $F : X \to X'$ is called *strongly monotone* if there exists an increasing bijection $\chi : [0, \infty) \to [0, \infty)$ with

$$\chi(\|u - v\|_X) \le \frac{\langle F(u) - F(v), u - v\rangle_X}{\|u - v\|_X}$$

for all $u, v \in X$.

We consider a conforming discretization of a strongly monotone problem in the following theorem.

Theorem 4.5 (Monotone problems) *Assume that $u \in X$ and $u_h \in X_h$ satisfy*

$$F(u)[v] = \ell(v), \qquad F(u_h)[v_h] = \ell(v_h)$$

for all $v \in X$ and $v_h \in X_h$, respectively, and let $\mathscr{C}_h(u)$ and $\mathscr{R}_h(u_h)$ for an interpolation operator i_h be defined by

$$\mathscr{C}_h(u; v_h) = F(i_h u)[v_h] - \ell(v_h), \qquad \mathscr{R}_h(u_h; v) = F(u_h)[v] - \ell(v)$$

for all $v_h \in X_h$ and $v \in X$, respectively. Then we have the a priori and a posteriori error estimates

$$\chi(\|i_h u - u_h\|_X) \le \|\mathscr{C}_h(u)\|_{X_h'}, \quad \chi(\|u - u_h\|_X) \le \|\mathscr{R}_h(u_h)\|_{X'}.$$

Proof We have

$$\|i_h u - u_h\|_X \, \chi(\|i_h u - u_h\|_X) \le \langle F(i_h u) - F(u_h), i_h u - u_h\rangle = \mathscr{C}_h(u; i_h u - u_h)$$

and

$$\|u - u_h\|_X \, \chi(\|u - u_h\|_X) \le \langle F(u) - F(u_h), u - u_h\rangle = -\mathscr{R}_h(u_h; u - u_h).$$

Dividing by $\|i_h u - u_h\|_X$ and $\|u - u_h\|_X$, respectively, yields the estimates. □

Example 4.13 (*p-Laplacian*) The p-Laplacian $-\operatorname{div}(|\nabla u|^{p-2}\nabla u)$ is identified with the functional $F : W_D^{1,p}(\Omega) \to W_D^{1,p}(\Omega)'$ defined by

$$F(u)[v] = \int_\Omega |\nabla u|^{p-2}\nabla u \cdot \nabla v \, dx$$

for $u, v \in W_D^{1,p}(\Omega)$. The functional F is the Fréchet derivative $F = DI$ of

$$I(u) = \frac{1}{p}\int_\Omega |\nabla u|^p \, dx.$$

If $p \ge 2$, then F is monotone with $\chi(s) = \alpha s^{p-1}$ for all $s \ge 0$ and some $\alpha > 0$. The functional is locally Lipschitz continuous in the sense that

$$\|F(u) - F(v)\|_{W_D^{1,p}(\Omega)'} \le M(\|\nabla u\|_{L^p(\Omega)} + \|\nabla v\|_{L^p(\Omega)})^{p-2}\|\nabla(u - v)\|_{L^p(\Omega)}$$

for a constant $M \in \mathbb{R}$ and $u, v \in W_D^{1,p}(\Omega)$. This estimate implies the consistency of conforming discretizations, e.g., with $\mathscr{S}_D^1(\mathscr{T}_h)$, and we obtain the error estimate

$$\alpha\|\nabla(i_h u - u_h)\|_{L^p(\Omega)}^{p-1} \le M\|\nabla(u - i_h u)\|_{L^p(\Omega)};$$

thus $\|\nabla(u - u_h)\|_{L^p(\Omega)} \le ch^{1/(p-1)}$ if $u \in W^{2,p}(\Omega) \cap W_D^{1,p}(\Omega)$.

If the operator F fails to be monotone but has a regular Fréchet derivative in the neighborhood of a solution, then a local error estimate follows from the implicit function theorem. For ease of presentation and without loss of generality, we consider the homogeneous problem $F(u) = 0$.

Theorem 4.6 (Local error estimate [10]) *Suppose that $F : X \to X'$ is continuous and $u \in X$ satisfies $F(u) = 0$. Assume that there exist constants $c_1, c_2, c_3, \varepsilon > 0$ with $c_2 < c_1$ such that*

$$\begin{aligned} \|F(u) - F(v)\|_{X'} &\le c_0 \|u - v\|_X, \\ \|DF(v)^{-1}\|_{L(X',X)} &\le c_1^{-1}, \\ \|DF(v) - DF(w)\|_{L(X,X')} &\le c_2 \|v - w\|_X \end{aligned}$$

for all $v, w \in B_\varepsilon(u)$. Let $i_h u \in X_h$ be an interpolant of u such that $c_0 \|i_h u - u\|_X \le (c_1 - c_2)\varepsilon$. Then there exists a unique $u_h \in X_h$ with $F(u_h) = 0$ and $\|u - u_h\|_X \le \varepsilon$.

Proof The assumptions of the theorem imply that

$$\|F(i_h u)\|_{X'} = \|F(i_h u) - F(u)\|_{X'} \le c_0 \|u - i_h u\|_X.$$

A quantitative version of the implicit function theorem, cf. [2], implies the existence of a unique $u_h \in X_h$ with the asserted properties. □

Example 4.14 (*Semilinear diffusion*) The theorem implies error estimates for the approximation of the semilinear equation

$$-\Delta u + f(u) = 0 \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega,$$

provided that f' and a solution $u \in H_0^1(\Omega)$ are such that the operator $-\Delta + f'(v)\mathrm{id}$ is invertible for all $v \in B_\varepsilon(u)$ for some $\varepsilon > 0$. It is sufficient for this that $f' > -c_P^{-2}$ with the smallest constant $c_P > 0$, such that $\|w\| \le c_P \|\nabla w\|$ for all $w \in H_0^1(\Omega)$.

The following proposition generalizes the Lax–Milgram and the Céa lemma to bilinear forms that are not elliptic.

Proposition 4.2 (Generalized Lax–Milgram and Céa lemma [1, 13]) *Let X, Y be Hilbert spaces, $a : X \times Y \to \mathbb{R}$ a continuous bilinear form with continuity constant M, and $\ell \in Y'$. Assume that there exists $\alpha > 0$ such that*

$$\sup_{v \in Y \setminus \{0\}} \frac{a(u, v)}{\|v\|_Y} \ge \alpha \|u\|_X$$

for all $u \in X$ and that for all $v \in Y \setminus \{0\}$, there exists $u \in Y$ with $a(u, v) \ne 0$. Then there exists a unique $u \in X$ with

$$a(u, v) = \ell(v)$$

for all $v \in Y$ and $\|u\|_X \le \alpha^{-1} \|\ell\|_{Y'}$. If $X_h \subset X$ and $Y_h \subset Y$ are such that the above conditions are satisfied with X and Y replaced by X_h and Y_h, respectively, then there *exists a unique* $u_h \in X_h$ *with*

$$a(u_h, v_h) = \ell(v_h)$$

for all $v_h \in Y_h$, *and we have*

$$\|u - u_h\|_X \le (1 + \alpha^{-1} M) \inf_{w_h \in X_h} \|u - w_h\|_X.$$

Proof Identifying the bilinear form a with the operator $A : X \to Y'$, we see that A is injective, i.e., $Au = 0$ for $u \in X$ implies $u = 0$. Noting that

$$\alpha \|u_j - u_k\|_X \le \sup_{v \in Y \setminus \{0\}} \frac{\langle A(u_j - u_k), v \rangle}{\|v\|_Y} \le \|Au_j - Au_k\|_{Y'}$$

proves that the range of A is closed. If $v \in Y$ is such that $\langle Au, v\rangle = 0$ for all $u \in X$, then the assumptions imply $v = 0$. Hence, the closed range theorem yields that the range of A is Y' and it follows that A is bijective, i.e., there exists a unique $u \in X$ with $Au = \ell$. The estimate for $\|u\|_X$ is an immediate consequence of the assumptions. The same arguments show that the operator $A_h : X_h \to Y_h'$ is an isomorphism and hence there exists a unique $u_h \in X_h$ with the asserted properties. Let $w_h \in X_h$, and for every $v_h \in X_h$ define

$$\widetilde{\ell}(v_h) = a(u - w_h, v_h).$$

Then there exists a unique $z_h \in X_h$ with $a(z_h, v_h) = \widetilde{\ell}(v_h)$ and $\|z_h\|_X \le \alpha^{-1} \|\widetilde{\ell}\|_{Y_h'}$. Since $a(u_h, v_h) = a(u, v_h)$ it follows that $z_h = u_h - w_h$, and hence

$$\|u_h - w_h\|_X \le \alpha^{-1} M \|u - w_h\|.$$

The triangle inequality implies the asserted estimate. □

Example 4.15 (*Helmholtz equation*) Let $\omega \in \mathbb{R}$ and $a : H_0^1(\Omega) \times H_0^1(\Omega) \to \mathbb{R}$ be for $u, v \in H_0^1(\Omega)$ defined by

$$a(u, v) = (\nabla u, \nabla v) - \omega^2 (u, v),$$

which corresponds to the partial differential equation $-\Delta u - \omega^2 u = f$ in Ω with boundary condition $u|_{\partial\Omega} = 0$. If ω^2 is not an eigenvalue of $-\Delta$, then a satisfies the conditions of the proposition. To prove this, note that $(-\Delta)^{-1} : L^2(\Omega) \to H_0^1(\Omega) \subset L^2(\Omega)$ is selfadjoint and compact with trivial kernel, so that there exists a complete orthonormal system $(u_j)_{j\in\mathbb{N}} \subset L^2(\Omega)$ of eigenfunctions of $(-\Delta)^{-1}$, i.e., for every $j \in \mathbb{N}$ we have $-\Delta u_j = \lambda_j u_j$ with positive eigenvalues $(\lambda_j)_{j\in\mathbb{N}}$ that do not accumulate at zero. We have $\lambda_j^{-1}(\nabla u_j, \nabla u_k) = (u_j, u_k) = \delta_{jk}$ for all $j, k \in \mathbb{N}$. Given $u = \sum_{j\in\mathbb{N}} \alpha_j u_j \in H_0^1(\Omega)$, define $v = \sum_{j\in\mathbb{N}} \sigma_j \alpha_j u_j$ with $\sigma_j = \operatorname{sign}(\|\nabla u_j\|^2 - \omega^2 \|u_j\|^2)$. Then

$$a(u, v) = \sum_{j\in\mathbb{N}} \sigma_j \alpha_j^2 \big(\|\nabla u_j\|^2 - \omega^2 \|u_j\|^2\big) \ge \min_{j\in\mathbb{N}} \frac{|\lambda_j - \omega^2|}{\lambda_j} \|\nabla u\|^2$$

and with $\|\nabla u\| = \|\nabla v\|$, we deduce that

$$\sup_{v \in H_0^1(\Omega)} \frac{a(u,v)}{\|\nabla v\|} \geq c_H \|\nabla u\|.$$

The second condition of the proposition is a direct consequence of the requirement that ω^2 is not an eigenvalue of $-\Delta$.

Remark 4.5 Proposition 4.2 is important for the analysis of saddle-point problems; the seminal paper [7] provides conditions that imply the assumptions of the proposition.

4.2.3 Abstract Subdifferential Flow

The subdifferential flow of a convex and lower semicontinuous functional $I : H \to \mathbb{R} \cup \{+\infty\}$ arises as an evolutionary model in applications, and can be used as a basis for numerical schemes to minimize I. The corresponding differential equation seeks $u : [0, T] \to H$, such that $u(0) = u_0$ and

$$\partial_t u \in -\partial I(u),$$

i.e., $u(0) = u_0$ and

$$(-\partial_t u, v - u)_H + I(u) \leq I(v)$$

for almost every $t \in [0, T]$ and every $v \in H$. An implicit discretization of this nonlinear evolution equation is equivalent to a sequence of minimization problems involving a quadratic term. We recall that $d_t u^k = (u^k - u^{k-1})/\tau$ denotes the backward difference quotient.

Theorem 4.7 (Semidiscrete scheme [15, 17]) *Assume that $I \geq 0$ and for $u^0 \in H$ let $(u^k)_{k=1,\dots,K} \subset H$ be minimizers for*

$$I_\tau^k(w) = \frac{1}{2\tau} \|w - u^{k-1}\|_H^2 + I(w)$$

for $k = 1, 2, \dots, K$. For $L = 1, 2, \dots, K$, we have

$$I(u^L) + \tau \sum_{k=1}^{L} \|d_t u^k\|_H^2 \leq I(u^0).$$

With the computable quantities

$$\mathscr{E}_k = -\tau \|d_t u^k\|_H^2 - I(u^k) + I(u^{k-1})$$

and the affine interpolant $\widehat{u}_\tau : [0,T] \to H$ *of the sequence* $(u^k)_{k=0,\dots,K}$ *we have the a posteriori error estimate*

$$\max_{t\in[0,T]} \|u - \widehat{u}\|_H^2 \le \|u_0 - u^0\|_H^2 + \tau \sum_{k=1}^{L} \mathscr{E}_k.$$

We have the a priori error estimate

$$\max_{k=0,\dots,K} \|u(t_k) - u^k\|_H^2 \le \|u_0 - u^0\|_H^2 + \tau I(u^0),$$

and under the condition $\partial I(u^0) \neq \emptyset$*, the improved variant*

$$\max_{k=0,\dots,K} \|u(t_k) - u^k\|_H^2 \le \|u_0 - u^0\|_H^2 + \tau^2 \|\partial^o I(u^0)\|_H^2,$$

where $\partial^o I(u^0) \in H$ *denotes the element of minimal norm in* $\partial I(u^0)$.

Proof The direct method in the calculus of variations yields that for $k = 1, 2, \dots, K$, there exists a unique minimizer $u^k \in H$ for I_τ^k, and we have $d_t u^k \in -\partial I(u^k)$, i.e.,

$$(-d_t u^k, v - u^k)_H + I(u^k) \le I(v)$$

for all $v \in H$; the choice of $v = u^{k-1}$ implies that

$$-\mathscr{E}_k = \tau \|d_t u^k\|_H^2 + I(u^k) - I(u^{k-1}) \le 0$$

with $0 \le \mathscr{E}_k \le -\tau d_t I(u^k)$. A summation over $k = 1, 2, \dots, L$ yields the asserted stability estimate. If $\widehat{u}_\tau$ is the piecewise affine interpolant of $(u^k)_{k=0,\dots,K}$ associated to the time steps $t_k = k\tau$, $k = 0, 1, \dots, K$, and u_τ^+ is such that $u_\tau^+|_{(t_{k-1},t_k)} = u^k$ for $k = 1, 2, \dots$ and $t_k = k\tau$, then we have

$$(-\partial_t \widehat{u}_\tau, v - u_\tau^+)_H + I(u_\tau^+) \le I(v)$$

for almost every $t \in [0, T]$ and all $v \in H$. In introducing

$$\mathscr{C}_\tau(t) = (-\partial_t \widehat{u}_\tau, u_\tau^+ - \widehat{u}_\tau)_H - I(u_\tau^+) + I(\widehat{u}_\tau)$$

we have

$$(-\partial_t \widehat{u}_\tau, v - \widehat{u}_\tau)_H + I(\widehat{u}_\tau) \le I(v) + \mathscr{C}_\tau(t).$$

The choice of $v = u$ in this inequality and $v = \widehat{u}_\tau$ in the continuous evolution equation yield

$$\frac{d}{dt}\frac{1}{2}\|u - \widehat{u}\|_H^2 = (-\partial_t[u - \widehat{u}_\tau], \widehat{u}_\tau - u)_H \le \mathscr{C}_\tau(t).$$

Noting $\widehat{u}_\tau - u_\tau^+ = (t - t_k)\partial_t \widehat{u}_\tau$ for $t \in (t_{k-1}, t_k)$ and using the convexity of I, i.e.,

$$I(\widehat{u}_\tau) \le \frac{t_k - t}{\tau} I(u^{k-1}) + \frac{t - t_{k-1}}{\tau} I(u^k),$$

we verify for $t \in (t_{k-1}, t_k)$ using $u_\tau^+ = u^k$ that

$$\mathscr{C}_\tau(t) \le (t - t_k)\|\partial_t \widehat{u}_\tau\|_H^2 - I(u_\tau^+) + \frac{t_k - t}{\tau} I(u^{k-1}) + \frac{t - t_{k-1}}{\tau} I(u^k) = \frac{t_k - t}{\tau}\mathscr{E}_k.$$

With $\mathscr{E}_k \le -\tau d_t I(u^k)$ and $I \ge 0$ we deduce that

$$\int_0^{t_L} \mathscr{C}_\tau(t)\, \mathrm{d}t \le \tau \sum_{k=1}^{L} \mathscr{E}_k \le -\tau^2 \sum_{k=1}^{L} d_t I(u^k) = -\tau\big(I(u^L) - I(u^0)\big) \le \tau I(u^0),$$

which implies the a posteriori and the first a priori error estimate. Assume that $\partial I(u^0) \neq \emptyset$ and define $u^{-1} \in H$ so that $d_t u^0 = (u^0 - u^{-1})/\tau = -\partial^o I(u^0)$, i.e., the discrete evolution equation also holds for $k = 0$,

$$(-d_t u^0, v - u^0)_H + I(u^0) \le I(v)$$

for all $v \in H$. Choosing $v = u^k$ in the equation for $d_t u^{k-1}$, $k = 1, 2, \ldots, K$, we observe that

$$(-d_t u^{k-1}, u^k - u^{k-1})_H + I(u^{k-1}) \le I(u^k),$$

i.e., $-\tau d_t I(u^k) \le \tau (d_t u^k, d_t u^{k-1})_H$, and it follows that

$$\begin{aligned}
\mathscr{E}_k &= -\tau(d_t u^k, d_t u^k)_H - \tau d_t I(u^k) \le -\tau(d_t u^k, d_t u^k)_H + \tau(d_t u^{k-1}, d_t u^k)_H \\
&= -\tau^2 (d_t^2 u^k, d_t u^k)_H = -\tau^2 \frac{d_t}{2}\|d_t u^k\|_H^2 - \frac{\tau^3}{2}\|d_t^2 u^k\|_H^2 \le -\tau^2 \frac{d_t}{2}\|d_t u^k\|_H^2.
\end{aligned}$$

This implies that

$$\int_0^{t_L} \mathscr{C}_\tau(t)\, \mathrm{d}t \le \tau \sum_{k=1}^{L} \mathscr{E}_k \le \frac{\tau^2}{2}\|d_t u^0\|_H^2 = \frac{\tau^2}{2}\|\partial^o I(u^0)\|_H^2,$$

which proves the improved a priori error estimate. □

Remarks 4.6 (i) The condition $\partial I(u^0) \neq \emptyset$ is restrictive in many applications.
(ii) Subdifferential flows $\partial_t u \in -\partial I(u)$, i.e., $Lu \ni 0$ for $Lu = \partial_t u + v$ with $v \in \partial I(u)$, and with a convex functional $I : H \to \mathbb{R} \cup \{+\infty\}$ define monotone

problems in the sense that

$$\begin{aligned}\big(Lu_1 - Lu_2, u_1 - u_2\big)_H &= (\partial_t(u_1 - u_2) + (v_1 - v_2), u_1 - u_2\big)_H \\ &\geq \big(\partial_t(u_1 - u_2), u_1 - u_2\big)_H = \frac{1}{2}\frac{d}{dt}\|u_1 - u_2\|_H^2\end{aligned}$$

for u_1, u_2 and v_1, v_2 with $v_i \in \partial I(u_i)$, $i = 1, 2$.
(iii) If $I : H \to \mathbb{R} \cup \{+\infty\}$ is strongly monotone in the sense that $(u_1 - u_2, v_1 - v_2)_H \geq \alpha\|u_1 - u_2\|_H^2$ whenever $v_\ell \in \partial I(u_\ell)$, $\ell = 1, 2$, and if there exists a solution $\overline{u} \in H$ of the stationary inclusion $\overline{v} = 0 \in \partial I(\overline{u})$, then we have $u(t) \to \overline{u}$ as $t \to \infty$. A proof follows from the estimate

$$\frac{1}{2}\frac{d}{dt}\|u - \overline{u}\|_H^2 = -(v - \overline{v}, u - \overline{u})_H \leq -\alpha\|u - \overline{u}\|_H^2,$$

where $v = -\partial_t u \in \partial I(u)$, and an application of Gronwall's lemma.

4.2.4 Weak Continuity Methods

Let $(u_h)_{h>0} \subset X$ be a bounded sequence in the reflexive, separable Banach space X such that there exists a weak limit $u \in X$ of a subsequence that is not relabeled, i.e., we have $u_h \rightharpoonup u$ as $h \to 0$. For an operator $F : X \to X'$, we define the sequence $(\xi_h)_{h>0} \subset X'$ through $\xi_h = F(u_h)$, and if the sequence is bounded in X', then there exists $\xi \in X'$, such that for a further subsequence $(\xi_h)_{h>0}$ which again is not relabeled, we have $\xi_h \rightharpoonup^* \xi$. The important question is now whether we have weak continuity in the sense that

$$F(u) = \xi.$$

Notice that weak continuity is a strictly stronger notion of continuity than strong continuity. For partial differential equations, this property is called *weak precompactness* of the solution set of the homogeneous equation, i.e., if $(u_j)_{j\in\mathbb{N}}$ is a sequence with $F(u_j) = 0$ for all $j \in \mathbb{N}$ and $u_j \rightharpoonup u$ as $j \to \infty$ then we may deduce that $F(u) = 0$. Such implications may also be regarded as properties of weak stability since they imply that if $F(u_j) = r_j$ with $\|r_j\|_{X'} \leq \varepsilon_j$ and $\varepsilon_j \to 0$ as $j \to \infty$, then we have $F(u) = 0$ for every accumulation point of the sequence $(u_j)_{j\in\mathbb{N}}$.

Theorem 4.8 (Discrete compactness) *For every $h > 0$ let $u_h \in X_h$ solve $F_h(u_h) = 0$. Assume that $F_h(u_h) \in X'$ with $\|F(u_h)\|_{X'} \leq c$ for all $h > 0$ and F is weakly continuous on X, i.e., $F(u_j)[v] \to F(u)[v]$ for all $v \in X$ whenever $u_j \rightharpoonup u$ in X. Suppose that for every bounded sequence $(w_h)_{h>0} \subset X$ with $w_h \in X_h$ for all $h > 0$, we have*

$$\|F(w_h) - F_h(w_h)\|_{X_h'} \to 0$$

as $h \to 0$ and $(X_h)_{h>0}$ is dense in X with respect to strong convergence. If $(u_h)_{h>0} \subset X$ is bounded, then there exists a subsequence $(u_{h'})_{h'>0}$ and $u \in X$ such that $u_h \rightharpoonup u$ in X and $F(u) = 0$.

Proof After extraction of a subsequence, we may assume that $u_h \rightharpoonup u$ in X as $h \to 0$ for some $u \in X$. Fixing $v \in X$ and using $F_h(u_h)[v_h] = 0$ for every $v_h \in X_h$, we have

$$F(u_h)[v] = F(u_h)[v - v_h] + F(u_h)[v_h] - F_h(u_h)[v_h].$$

For a sequence $(v_h)_{h>0} \subset X$ with $v_h \in X_h$ for every $h > 0$ and $v_h \to v$ in X, we find that

$$|F(u_h)[v - v_h]| \le \|F(u_h)\|_{X'} \|v - v_h\|_X \to 0$$

as $h \to 0$. The sequences $(u_h)_{h>0}$ and $(v_h)_{h>0}$ are bounded in X and thus

$$|F(u_h)[v_h] - F_h(u_h)[v_h]| \le \|F(u_h) - F_h(u_h)\|_{X_h'} \|v_h\|_X \to 0$$

as $h \to 0$. Together with the weak continuity of F we find that

$$F(u)[v] = \lim_{h\to 0} F(u_h)[v] = 0.$$

Since $v \in X$ was arbitrary this proves the theorem. $\square$

The crucial part in the theorem is the weak continuity of the operator F. We include an example of an operator related to a constrained nonlinear partial differential equation that fulfills this requirement.

Example 4.16 (*Harmonic maps*) Let $(u_j)_{j\in\mathbb{N}} \subset H^1(\Omega;\mathbb{R}^3)$ be a bounded sequence such that $|u_j(x)| = 1$ for all $j \in \mathbb{N}$ and almost every $x \in \Omega$. Assume that for every $j \in \mathbb{N}$ and all $v \in H^1(\Omega;\mathbb{R}^3) \cap L^\infty(\Omega;\mathbb{R}^3)$, we have

$$F(u_j)[v] = \int_\Omega \nabla u_j \cdot \nabla v \,\mathrm{d}x - \int_\Omega |\nabla u_j|^2 u_j \cdot v \,\mathrm{d}x = 0.$$

The choice of $v = u_j \times w$ shows that we have

$$\widetilde{F}(u_j)[w] = \int_\Omega \nabla u_j \cdot \nabla(u_j \times w) \,\mathrm{d}x = 0$$

for all $w \in H^1(\Omega;\mathbb{R}^3) \cap L^\infty(\Omega;\mathbb{R}^3)$. Using $\partial_k u_j \cdot \partial_k(u_j \times w) = \partial_k u_j \cdot (u_j \times \partial_k w)$ for $k = 1, 2, \dots, d$, we find that

$$\widetilde{F}(u_j)[w] = \sum_{k=1}^d \int_\Omega \partial_k u_j \cdot (u_j \times \partial_k w) \,\mathrm{d}x = 0.$$

If $u_j \rightharpoonup u$ in $H^1_{\mathrm{D}}(\Omega;\mathbb{R}^3)$, then $u_j \to u$ in $L^2(\Omega;\mathbb{R}^3)$ and thus, for every fixed $w \in C^\infty(\overline{\Omega};\mathbb{R}^3)$, we can pass to the limit and find that

$$\widetilde{F}(u)[w] = 0.$$

Since up to a subsequence we have $u_j(x) \to u(x)$ for almost every $x \in \Omega$, we verify that $|u(x)| = 1$ for almost every $x \in \Omega$. A density result shows that this holds for all $w \in H^1(\Omega;\mathbb{R}^3) \cap L^\infty(\Omega;\mathbb{R}^3)$. Reversing the above argument by choosing $w = u \times v$ and employing the identity $a \times (b \times c) = (b \cdot a)c - (c \cdot a)b$ shows that $F(u)[v] = 0$ for all $v \in H^1(\Omega;\mathbb{R}^3) \cap L^\infty(\Omega;\mathbb{R}^3)$.

A general concept for weak continuity is based on the notion of pseudomonotonicity.

Example 4.17 (*Pseudomonotone operators*) The operator $F : X \to X'$ is a *pseudomonotone* operator if it is bounded, i.e., $\|F(u)\|_{X'} \le c(1 + \|u\|_X^s)$ for some $s \ge 0$, and whenever $u_j \rightharpoonup u$ in X, we have the implication that

$$\limsup_{j\to\infty} F(u_j)[u_j - u] \le 0 \quad \Longrightarrow \quad F(u)[u - v] \le \liminf_{j\to\infty} F(u_j)[u_j - v].$$

For such an operator we have that if $F(u_h)[v_h] = \ell(v_h)$ for all $v_h \in X_h$ with a strongly dense family of subspaces $(X_h)_{h>0}$ and $u_h \rightharpoonup u$ as $h \to 0$, then $F(u) = \ell$. To verify this, let $v \in X$ and $(v_h)_{h>0}$ with $v_h \in X_h$ such that $v_h \to u$ and note that

$$\begin{aligned}\limsup_{h\to 0} F(u_h)[u_h - u] &= \limsup_{h\to 0} F(u_h)[u_h - v_h] + F(u_h)[v_h - u]\\ &= \limsup_{h\to 0} \ell(u_h - v_h) + F(u_h)[v_h - u] = 0.\end{aligned}$$

Pseudomonotonicity yields for every $v_{h'} \in \cup_{h>0} X_h$ that

$$F(u)[u - v_{h'}] \le \liminf_{h\to 0} F(u_h)[u_h - v_{h'}] = \lim_{h\to 0} \ell(u_h - v_{h'}) = \ell(u - v_{h'}).$$

With the density of $(X_h)_{h>0}$ in X, we conclude that $F(u)[u - v] \le \ell(u - v)$ for all $v \in X$ and with $v = u \pm w$, we find that $F(u)[w] = \ell(w)$ for all $w \in X$.

Remarks 4.7 (i) Radially continuous bounded operators are pseudomonotone. Here, radial continuity means that $t \mapsto F(u + tv)[v]$ is continuous for $t \in \mathbb{R}$ and all $u, v \in X$. These operators allow us to apply Minty's trick to deduce from the inequality $\ell(u - v) - F(v)[u - v] \ge 0$ for all $v \in X$ that $F(u) = \ell$. To prove this implication, note that with $v = u + \varepsilon w$, we find that $\ell(w) - F(u + \varepsilon w)[w] \le 0$ and by radial continuity for $\varepsilon \to 0$, it follows that $\ell(w) - F(u)[w] \le 0$ and hence $F(u) = \ell$.
(ii) Pseudomonotone operators are often of the form $F = F_1 + F_2$ with a monotone operator F_1 and a weakly continuous operator F_2, e.g., a lower-order term described by F_2.

Example 4.18 (*Quasilinear diffusion*) The concept of pseudomonotonicity applies to the quasilinear elliptic equation

$$-\operatorname{div}\left(|\nabla u|^{p-2}\nabla u\right)+g(u)=f \text{ in } \Omega, \quad u|_{\partial\Omega}=0 \text{ on } \partial\Omega,$$

with $g\in C(\mathbb{R})$ such that $|g(s)|\le c(1+|s|^{r-1})$ and $1<p<d, r<dp/(d-p)$.

4.3 Solution of Discrete Problems

We discuss in this section the practical solution of discretized minimization problems of the form

$$\text{Minimize } I_h(u_h)=\int_\Omega W(\nabla u_h)+g(u_h)\,\mathrm{d}x \text{ among } u_h\in\mathscr{A}_h.$$

In particular, we investigate four model situations with smooth and nonsmooth integrands and smooth and nonsmooth constraints included in $\mathscr{A}$. The iterative algorithms are based on an approximate solution of the discrete Euler–Lagrange equations. More general results can be found in the textbooks [4, 12].

4.3.1 Smooth, Unconstrained Minimization

Suppose that

$$\mathscr{A}_h=\{u_h\in\mathscr{S}^1(\mathscr{T}_h)^m : u_h|_{\Gamma_\mathrm{D}}=u_{\mathrm{D},h}\}$$

and I_h is defined as above with functions $W\in C^1(\mathbb{R}^{m\times d})$ and $g\in C^1(\mathbb{R}^m)$. The case $\Gamma_\mathrm{D}=\emptyset$ is not generally excluded in the following. A necessary condition for a minimizer $u_h\in\mathscr{A}_h$ is that for all $v_h\in\mathscr{S}^1_\mathrm{D}(\mathscr{T}_h)^m$, we have

$$F_h(u_h)[v_h]=\int_\Omega DW(\nabla u_h)\cdot\nabla v_h+Dg(u_h)\cdot v_h\,\mathrm{d}x=0.$$

Steepest descent methods successively lower the energy by minimizing in descent directions defined through an appropriate gradient.

Algorithm 4.1 (*Descent method*) *Let* $(\cdot,\cdot)_H$ *be a scalar product on* $\mathscr{S}^1_\mathrm{D}(\mathscr{T}_h)^m$ *and* $\mu\in(0,1/2)$. *Given* $u_h^0\in\mathscr{A}_h$, *compute the sequence* $(u_h^j)_{j=0,1,\ldots}$ *via* $u_h^{j+1}=u_h^j+\alpha_j d_h^j$ *with* $d_h^j\in\mathscr{S}^1_\mathrm{D}(\mathscr{T}_h)^m$ *such that*

$$(d_h^j,v_h)_H=-F_h(u_h^j)[v_h]$$

for all $v_h \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)^m$ *and either the fixed step-size*

$$\alpha_j = \tau$$

or the line-search minimum which seeks the maximal $\alpha_j \in \{2^{-\ell},\ \ell \in \mathbb{N}_0\}$ *such that*

$$I_h(u_h^j + \alpha_j d_h^j) \le I_h(u_h^j) - \mu\alpha_j \|d_h^j\|_H^2.$$

Stop the iteration if $\|\alpha_j d_h^j\|_H \le \varepsilon_{\mathrm{stop}}$.

Remarks 4.8 (i) Since I_h is continuously differentiable, the descent method decreases the energy in every step. This follows from

$$\frac{d}{d\alpha} I_h(u_h^j + \alpha d_h^j)\Big|_{\alpha=0} = DI_h(u_h^j)[d_h^j] = F_h(u_h^j)[d_h^j] = -\|d_h^j\|_H^2,$$

i.e., the continuous function $\varphi(\alpha) = I_h(u_h^j + \alpha d_h^j)$ is strictly decreasing for $\alpha \in [0,\delta]$. The existence of $\alpha_j > 0$ that satisfies the Armijo–Goldstein condition of Algorithm 4.1 follows from expanding

$$I_h(u_h^j + \alpha d_h^j) = I_h(u_h^j) - \alpha\|d_h^j\|_H^2 + \mathscr{O}(\alpha^2)$$

provided that W and g are sufficiently smooth so that $I_h \in C^2(X_h)$.
(ii) The scalar product $(\cdot,\cdot)_H$ acts like a preconditioner for F_h, i.e., we have $u_h^{j+1} = u_h^j - \tau X_H^{-1} F_h(u_h^j)$ with respect to an appropriate basis. In particular, the descent method may be regarded as a fixed-point iteration.
(iii) Larger step sizes are typically possible for implicit or semi-implicit versions of the descent method, i.e., by considering a fixed step-size and the modified equation

$$(d_h^j, v_h)_H + \widetilde{F}_h(u_h^j + \tau d_h^j, u_h^j)[v_h] = 0$$

for all $v_h \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)^m$ and with a function $\widetilde{F}_h$ such that $\widetilde{F}_h(u_h,u_h) = F_h(u_h)$. If $F_h(u_h) = G_h(u_h) + T_h(u_h)$ with a linear or monotone operator G_h, then a natural choice is $\widetilde{F}_h(u_h,\widetilde{u}_h) = G_h(u_h) + T_h(\widetilde{u}_h)$. Generally, large time steps are possible when monotone terms are treated implicitly and antimonotone terms explicitly.
(iv) If $X_h = V_h \times W_h$ and $I_h(u_h) = J_h(\phi_h,\psi_h)$ is separately convex, i.e., the mappings $v_h \mapsto J_h(v_h,\psi_h)$ and $w_h \mapsto J_h(\phi_h,w_h)$ are convex for all $(\phi_h,\psi_h) \in V_h \times W_h$, a decoupled, semi-implicit gradient flow discretization is unconditionally stable. Given the initial $(\phi_h^0,\psi_h^0) \in V_h \times W_h$, consider the iteration

$$\begin{aligned}(d_t\phi_h^{j+1}, v_h)_{V_h} + \delta_1 J_h(\phi_h^{j+1},\psi_h^j)[v_h] &= 0,\\ (d_t\psi_h^{j+1}, w_h)_{W_h} + \delta_2 J_h(\phi_h^{j+1},\psi_h^{j+1})[w_h] &= 0,\end{aligned}$$

where $\delta_1 J_h$ and $\delta_2 J_h$ denote the Fréchet derivatives of J_h with respect to the first and second argument, respectively. The choices $v_h = d_t \phi_h^{j+1}$, $w_h = d_t \psi_h^{j+1}$ and the separate convexity of J lead to

$$\begin{aligned}
\|d_t \phi_h^{j+1}\|_{V_h}^2 + \|d_t \psi_h^{j+1}\|_{W_h}^2 &= -\delta_1 J_h(\phi_h^{j+1}, \psi_h^j)[d_t \phi_h^{j+1}] \\
&\quad - \delta_2 J_h(\phi_h^{j+1}, \psi_h^{j+1})[d_t \psi_h^{j+1}] \\
&\leq \tau^{-1}\big(J_h(\phi_h^j, \psi_h^j) - J_h(\phi_h^{j+1}, \psi_h^j)\big) \\
&\quad + \tau^{-1}\big(J_h(\phi_h^{j+1}, \psi_h^j) - J_h(\phi_h^{j+1}, \psi_h^{j+1})\big) \\
&= -d_t J_h(\phi_h^{j+1}, \psi_h^{j+1}),
\end{aligned}$$

which implies the unconditional stability of the scheme.

Theorem 4.9 (Convex functionals) *Assume that I_h is convex and bounded from below and F_h is Lipschitz continuous, i.e., there exists $c_F \geq 0$ such that*

$$\|F_h(w_h) - F_h(v_h)\|_{X_h'} \leq c_F \|w_h - v_h\|_X$$

for all $w_h, v_h \in X_h$. Let $c_h > 0$ be such that $\|v_h\|_X \leq c_h \|v_h\|_H$ for all $v_h \in X_h$. Then the steepest descent method with fixed step-size $\tau > 0$ such that $\tau c_F c_h \leq 1/2$ terminates within a finite number of iterations, and for all $J \geq 0$, we have

$$I_h(u_h^{J+1}) + (\tau/2) \sum_{j=0}^{J} \|d_h^j\|_H^2 \leq I_h(u_h^0).$$

Proof The convexity of I_h implies that

$$F_h(u_h^{j+1})[u_h^{j+1} - u_h^j] + I_h(u_h^j) \geq I_h(u_h^{j+1}).$$

Using that $\tau d_h^j = u_h^{j+1} - u_h^j$ and choosing $v_h = \tau d_h^j$ in the discrete scheme leads to

$$\begin{aligned}
I_h(u_h^{j+1}) - I_h(u_h^j) + \tau \|d_h^j\|_H^2 &\leq (d_h^j, d_h^j)_H + \tau F_h(u_h^{j+1})[d_h^j] \\
&= (d_h^j, d_h^j)_H - F_h(u_h^j)[d_h^j] \\
&\quad + \tau\big(F_h(u_h^j) - F_h(u_h^{j+1})\big)[d_h^j] \\
&= \tau\big(F_h(u_h^j) - F_h(u_h^{j+1})\big)[d_h^j] \leq c_F c_h \tau^2 \|d_h^j\|_H^2.
\end{aligned}$$

Therefore, if $\tau c_F c_h \leq 1/2$ we deduce the estimate from a summation over $j = 0, 1, \ldots, J$. The estimate implies that $d_h^j \to 0$ as $j \to \infty$ so that $\|\tau d_h^j\|_H \leq \varepsilon_{\text{stop}}$ for j sufficiently large. $\square$

Remarks 4.9 (i) The arguments of the proof of the theorem show that the implicit version of the descent method, defined by $(d_h^j, v_h)_H + F_h(u_h^j + \tau d_h^j)[v_h] = 0$ for every $v_h \in \mathscr{S}_{\mathrm{D}}^1(\mathscr{T}_h)$, is unconditionally convergent, but requires the solution of nonlinear systems of equations in every time step.
(ii) For nonconvex functionals, the iteration typically converges to a local minimum of I_h. Theoretically, the iteration may stop at a saddle point or local maximum.

To formulate the Newton method for solving the equation $F_h(u_h) = 0$ in X_h' we assume that $W \in C^2(\mathbb{R}^{m \times d})$ and $g \in C^2(\mathbb{R}^m)$. The Newton scheme may be regarded as an explicit descent method with a variable metric defined by the second variation of the energy functional I_h, i.e.,

$$DF_h(u_h)[w_h, v_h] = \int_\Omega D^2 W(\nabla u_h)[\nabla w_h, \nabla v_h] + D^2 g(u_h)[w_h, v_h] \,\mathrm{d}x$$

for $u_h, v_h, w_h \in \mathscr{S}_{\mathrm{D}}^1(\mathscr{T}_h)^m$.

Algorithm 4.2 (*Newton method*) *Given $u_h^0 \in \mathscr{A}_h$, compute the sequence $(u_h^j)_{j=0,1,\ldots}$ via $u_h^{j+1} = u_h^j + \alpha_j d_h^j$ with $d_h^j \in \mathscr{S}_{\mathrm{D}}^1(\mathscr{T}_h)^m$ such that*

$$DF_h(u_h^j)[d_h^j, v_h] = -F_h(u_h^j)[v_h]$$

for all $v_h \in \mathscr{S}_{\mathrm{D}}^1(\mathscr{T}_h)^m$ and $\alpha_j > 0$ with either the optimal step-size $\alpha_j = 1$, a fixed damping parameter $\alpha_j = \tau < 1$, or a line search minimum α_j as in Algorithm 4.1. Stop the iteration if $\|\alpha_j d_h^j\|_H \le \varepsilon_{\mathrm{stop}}$ for a norm $\|\cdot\|_H$ on $\mathscr{S}_{\mathrm{D}}^1(\mathscr{T}_h)^m$.

The convergence of the Newton iteration will be discussed in a more general context below in Sect. 4.3.3.

Remark 4.10 As opposed to the descent method, the Newton iteration can in general only be expected to converge locally. Under certain conditions the Newton scheme converges quadratically in a neighborhood of a solution. Optimal results can be obtained by combining the globally but slowly convergent descent method with the locally but rapidly convergent Newton method. Since the convergence of the Newton method is often difficult to establish and requires W and g to be sufficiently regular, developing globally convergent schemes is important to construct reliable numerical methods.

Example 4.19 For the approximation of minimal surfaces that are presented by graphs of functions over Ω, we consider

$$I_h(u_h) = \int_\Omega (1 + |\nabla u_h|^2)^{1/2} \,\mathrm{d}x$$

and note that for $u_h \in \mathscr{A}_h = \{v_h \in \mathscr{S}^1(\mathscr{T}_h) : v_h|_{\Gamma_D} = u_{D,h}\}$ and $v_h, w_h \in \mathscr{S}^1_D(\mathscr{T}_h)$ we have

$$F_h(u_h)[v_h] = \int_\Omega \frac{\nabla u_h \cdot \nabla v_h}{(1 + |\nabla u_h|^2)^{1/2}} \,\mathrm{d}x$$

and

$$DF_h(u_h)[w_h, v_h] = \int_\Omega \frac{\nabla w_h \cdot \nabla v_h}{(1 + |\nabla u_h|^2)^{1/2}} - \frac{\big(\nabla u_h \cdot \nabla v_h\big)\big(\nabla u_h \cdot \nabla w_h\big)}{(1 + |\nabla u_h|^2)^{3/2}} \,\mathrm{d}x.$$

Figure 4.2 displays a combined MATLAB implementation of the Newton iteration and the descent method with line search. The Newton method fails to provide meaningful approximations for moderate perturbations of the nodal interpolant of the exact solution as a starting value.

4.3.2 Smooth Constrained Minimization

We next consider the case that the set of admissible functions includes a pointwise constraint, which is imposed at the nodes of a triangulation, i.e., for $G \in C(\mathbb{R}^m)$, we have

$$\mathscr{A}_h = \{u_h \in \mathscr{S}^1(\mathscr{T}_h)^m : u_h|_{\Gamma_D} = u_{D,h},\ G\big(u_h(z)\big) = 0 \quad \text{for all } z \in \mathscr{N}_h\}.$$

The identity $G\big(u_h(z)\big) = 0$ for all $z \in \mathscr{N}_h$ is equivalent to the condition $\mathscr{I}_h G(u_h) = 0$. We always assume in the following that $\mathscr{A}_h \neq \emptyset$, i.e., that the function $u_{D,h}$ is compatible with the constraint. Moreover, we assume $G \in C^1(\mathbb{R}^m)$ with $DG(s) \neq 0$ for every $s \in M = G^{-1}(\{0\})$ so that $M \subset \mathbb{R}^m$ is an $(m-1)$-dimensional C^1-submanifold. The Euler–Lagrange equations of the discrete minimization problem

$$I_h(u_h) = \int_\Omega W(\nabla u_h) \,\mathrm{d}x$$

in the set of all functions $u_h \in \mathscr{A}_h$ can then be formulated as follows.

Proposition 4.3 (Optimality conditions) *The function $u_h \in \mathscr{A}_h$ is stationary for I_h in $\mathscr{A}_h$ if and only if*

$$F_h(u_h)[w_h] = 0$$

for all $w_h \in T_{u_h}\mathscr{A}_h$, where the discrete tangent space $T_{u_h}\mathscr{A}_h$ of $\mathscr{A}_h$ at u_h is defined by

$$T_{u_h}\mathscr{A}_h = \{w_h \in \mathscr{S}^1_D(\mathscr{T}_h)^m : DG(u_h(z))w_h(z) = 0 \quad \text{forall } z \in \mathscr{N}_h \setminus \Gamma_D\}.$$

```
function min_surf(red,scheme)
[c4n,n4e,Db,Nb] = triang_ring(red); nC = size(c4n,1);
[s,m,m_lumped] = fe_matrices(c4n,n4e); X_metric = s;
dNodes = unique(Db); fNodes = setdiff(1:nC,dNodes);
u = u_D(c4n); sd = zeros(nC,1);
pert = .01*(rand(nC,1)-.5); pert(dNodes) = 0; u = u+pert;
mu = 1/4; norm_corr = 1; eps_stop = 1e-4;
while norm_corr > eps_stop
    alpha = 1;
    if strcmp(scheme,'descent')
        [I_0,dI,¬] = energy(c4n,n4e,u);
        sd(fNodes) = -X_metric(fNodes,fNodes)\dI(fNodes);
        [I_alpha,¬] = energy(c4n,n4e,u+alpha*sd);
        armijo = I_alpha-I_0-mu*alpha*dI'*sd;
        while armijo > 0
            alpha = alpha/2;
            [I_alpha,¬] = energy(c4n,n4e,u+alpha*sd);
            armijo = I_alpha-I_0-mu*alpha*dI'*sd;
        end
    elseif strcmp(scheme,'newton')
        [¬,dI,d2I] = energy(c4n,n4e,u);
        sd(fNodes) = -d2I(fNodes,fNodes)\dI(fNodes);
    end
    u = u+alpha*sd; show_p1(c4n,n4e,Db,Nb,u);
    norm_corr = sqrt((alpha*sd)'*X_metric*(alpha*sd))
end

function [I,dI,d2I] = energy(c4n,n4e,u)
[nC,d] = size(c4n); nE = size(n4e,1);
ctr_max = (d+1)^2*nE; ctr = 0;
I1 = zeros(ctr_max,1); I2 = zeros(ctr_max,1);
X_d2I = zeros(ctr_max,1);
I = 0; dI = zeros(nC,1);
for j = 1:nE
    grads_T = [1,1,1;c4n(n4e(j,:),:)']\[0,0;eye(2)];
    vol_T = det([1,1,1;c4n(n4e(j,:),:)'])/2;
    du = (grads_T)'*u(n4e(j,:)); mod_du = norm(du);
    I = I+vol_T*(1+mod_du^2)^(1/2);
    P_loc = ((1+mod_du^2).*eye(d)-du*du')./((1+mod_du^2)^(3/2));
    for k = 1:d+1
        dI(n4e(j,k)) = dI(n4e(j,k))...
            +vol_T*grads_T(k,:)*du/(1+mod_du^2);
        for ell = 1:d+1
            ctr = ctr+1; I1(ctr) = n4e(j,k); I2(ctr) = n4e(j,ell);
            X_d2I(ctr) = vol_T*(P_loc*grads_T(k,:)')'...
                *grads_T(ell,:)';
        end
    end
end
d2I = sparse(I1,I2,X_d2I);

function val = u_D(x)
val = .5*(2-sqrt(sum(x.^2,2)));
```

Fig. 4.2 MATLAB routine for the computation of discrete minimal surfaces with the Newton and the steepest descent scheme

Proof We let $\varphi_h : (-\varepsilon, \varepsilon) \to \mathscr{A}_h$ be a continuously differentiable function with $\varphi_h(0) = u_h$. We then have that $w_h = \varphi_h'(0) \in T_{u_h}\mathscr{A}_h$ and

$$0 = \frac{d}{dt} I_h(\varphi_h(t))\Big|_{t=0} = DI_h(u_h)[w_h].$$

Conversely, for every $w_h \in T_{u_h}\mathscr{A}_h$ there exists a function $\varphi_h(t)$ as above. □

Remark 4.11 An equivalent characterization of stationary points is the existence of a Lagrange multiplier $\lambda_h \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)$ such that for all $v_h \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)^m$, we have

$$F_h(u_h)[v_h] + (\lambda_h DG(u_h), v_h)_h = 0.$$

We propose the following descent scheme for the iterative solution of the constrained problem. It may be regarded as a semi-implicit discretization of an H-gradient flow. In particular, the problems that have to be solved at every step of the iteration are linear if F_h is linear.

Algorithm 4.3 (*Constrained descent method*) *Let* $(\cdot,\cdot)_H$ *be a scalar product on* $\mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)^m$ *and given* $u_h^0 \in \mathscr{A}_h$, *compute the sequence* $(u_h^j)_{j=0,1,\ldots}$ *via* $u_h^{j+1} = u_h^j + \tau d_h^j$ *with* $d_h^j \in T_{u_h^j}\mathscr{A}_h$ *such that*

$$(d_h^j, v_h)_H + F_h(u_h^j + \tau d_h^j)[v_h] = 0$$

for all $v_h \in T_{u_h^j}\mathscr{A}_h$. *Stop the iteration if* $\|d_h^j\|_H \le \varepsilon_{\mathrm{stop}}$.

Remark 4.12 If F_h is linear, then the solution of an iteration is equivalent to the solution of a linear system of equations of the form

$$\begin{bmatrix} X_H + \tau S & dG^\top \\ dG & 0 \end{bmatrix} \begin{bmatrix} D_h^j \\ \Lambda_h^j \end{bmatrix} = \begin{bmatrix} -SU_h^j \\ 0 \end{bmatrix},$$

where D_h^j, U_h^j, and Λ_h^j are vectors that contain the nodal values of the functions d_h^j, u_h^j, and λ_h^j, respectively, and X_H, S, and dG_h are matrices that represent the scalar product $(\cdot,\cdot)_H$, the bilinear form $F_h(u_h)[v_h]$, and the linearized constraint defined by DG.

The iterates $(u_h^j)_{j=0,1,\ldots}$ will in general not satisfy the constraint $\mathscr{I}_h G(u_h^j) = 0$ but under moderate conditions, the violation of the constraint is small. We recall the notation $\|v\|_h^2 = \int_\Omega \mathscr{I}_h[v^2]\,\mathrm{d}x$ for $v \in C(\overline{\Omega})$.

Theorem 4.10 (Constrained convex minimization) *Assume that* $G \in C^2(\mathbb{R}^m)$ *with* $\|D^2G\|_{L^\infty(\mathbb{R}^m)} \le c$, I_h *is convex,* $u_h^0 \in \mathscr{A}_h$, *and* $\|v_h\|_h \le c\|v_h\|_H$ *for all* $v_h \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)^m$. *For all* $J \ge 0$ *we have*

$$I_h(u_h^{J+1}) + \tau \sum_{j=0}^{J} \|d_h^j\|_H^2 \le I_h(u_h^0),$$

and for every $j = 1, 2, \ldots,$ the bound

$$\|\mathscr{I}_h G(u_h^{j+1})\|_{L^1(\Omega)} \le c\tau I_h(u_h^0).$$

The algorithm terminates after a finite number of iterations.

Proof The convexity of I_h implies that

$$I_h(u_h^j + \tau d_h^j) + F_h(u_h + \tau d_h^j)[u_h^j - (u_h^j + \tau d_h^j)] \le I_h(u_h^j).$$

With the choice of $v_h = \tau d_h^j$ in the algorithm and the relation $u_h^{j+1} = u_h^j + \tau d_h^j$, this leads to

$$I_h(u_h^{j+1}) - I_h(u_h^j) \le \tau F_h(u_h^{j+1})[d_h^j] = -\tau \|d_h^j\|_H^2.$$

A summation over $j = 0, 1, \ldots, J$ proves the energy law. A Taylor expansion shows that for every $z \in \mathscr{N}_h \setminus \Gamma_{\mathrm{D}}$, we have for some $\xi_z^j \in \mathbb{R}^m$ that

$$G(u_h^{j+1}(z)) = G(u_h^j(z)) + \tau DG(u_h^j(z)) \cdot d_h^j(z) + \frac{\tau^2}{2} d_h^j(z)^\top D^2 G(\xi_z^j) d_h^j(z).$$

Noting $DG(u_h^j(z)) \cdot d_h^j(z) = 0$ and $G(u_h^0(z)) = 0$, we deduce by induction that

$$G(u_h^{J+1}(z)) = \frac{\tau^2}{2} \sum_{j=0}^{J} d_h^j(z)^\top D^2 G(\xi_z^j) d_h^j(z).$$

Since $D^2 G$ is uniformly bonded we have with $\beta_z = \int_\Omega \varphi_z \, \mathrm{d}x$ that

$$\begin{aligned}
\|\mathscr{I}_h G(u_h^{j+1})\|_{L^1(\Omega)} &\le \sum_{z \in \mathscr{N}} \beta_z |G(u_h^j(z))| \le c\tau^2 \sum_{j=0}^{J} \sum_{z \in \mathscr{N}} \beta_z |d_h^j(z)|^2 \\
&= c\tau^2 \sum_{j=0}^{J} \|d_h^j\|_h^2.
\end{aligned}$$

A combination with the energy law implies the bound for $\|\mathscr{I}_h G(u_h^{j+1})\|_{L^1(\Omega)}$. The convergence of the iteration follows from the convergence of the sum of norms of the correction vectors d_h^j. □

Remark 4.13 In order to satisfy the constraint exactly, the algorithm can be augmented by defining the new iterates through the projection

$$u_h^{j+1}(z) = \pi_M\big(u_h^j(z) + \tau d_h^j(z)\big),$$

where $\pi_M : U_\delta(M) \to M$ is the nearest neighbor projection onto $M = G^{-1}(\{0\})$ that is defined in a tubular neighborhood $U_\delta(M)$ of M for some $\delta > 0$ if $M \in C^2$. The step-size $\tau > 0$ has to be sufficiently small in order to guarantee the well-posedness of the iteration.

Example 4.20 (*Harmonic maps*) Minimizing the Dirichlet energy in the set

$$\mathscr{A}_h = \{u_h \in \mathscr{S}^1(\mathscr{T}_h)^m : u_h|_{\Gamma_{\mathrm{D}}} = u_{\mathrm{D},h},\ |u_h(z)| = 1 \quad \text{for all } z \in \mathscr{N}_h\}$$

corresponds to the situation of Theorem 4.10 with $G(s) = |s|^2 - 1$ and $M = S^{m-1} = \{s \in \mathbb{R}^m : |s| = 1\}$. In particular, we have $DG(s) = 2s$ and $\|D^2 G\|_{L^\infty(\mathbb{R}^m)} = 2m^{1/2}$. The discrete tangent spaces are given by

$$T_{u_h}\mathscr{A}_h = \{w_h \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)^m : u_h(z) \cdot w_h(z) = 0 \quad \text{for all } z \in \mathscr{N}_h \setminus \Gamma_{\mathrm{D}}\}.$$

The nearest neighbor projection π_{S^2} is for $s \in \mathbb{R}^m \setminus \{0\}$ defined by $\pi_{S^2}(s) = s/|s|$.

4.3.3 Nonsmooth Equations

We consider an abstract equation of the form

$$F_h(u_h)[v_h] = 0$$

for all $v_h \in X_h$ with a continuous operator $F_h : X_h \to Y_h$ that may not be continuously differentiable. The goal is to formulate conditions that allow us to prove convergence of an appropriate generalization of the Newton method. We let X_h and Y_h be Banach spaces in the following, and assume that X_h is equipped with the norm of a Banach space X. We let $\mathrm{L}(X_h, Y_h)$ denote the space of continuous linear operators $A_h : X_h \to Y_h$ and let $\|A_h\|_{\mathrm{L}(X_h, Y_H)}$ be the corresponding operator norm.

Definition 4.3 We say that $F_h : X_h \to Y_h$ is *Newton differentiable* at $v_h \in X_h$ if there exists $\varepsilon > 0$ and a function $G_h : B_\varepsilon(v_h) \to \mathrm{L}(X_h, Y_h)$ such that

$$\lim_{w_h \to 0} \frac{\|F_h(v_h + w_h) - F_h(v_h) - G_h(v_h + w_h)[w_h]\|_{Y_h}}{\|w_h\|_X} = 0.$$

The function G_h is called the *Newton derivative* of F_h at v_h.

Remark 4.14 Notice that in contrast to the definition of the classical derivative, here the derivative is evaluated at the perturbed point $v_h + w_h$. This is precisely the expression that arises in the convergence analysis of the classical Newton iteration.

Examples 4.21 (i) If $F_h : X_h \to Y_h$ is continuously differentiable in a neighborhood of $v_h \in X_h$, then F_h is Newton differentiable at v_h with Newton derivative

$G_h = DF_h$, i.e., we have

$$\begin{aligned}\big\|F_h(v_h + w_h) - F_h(v_h) - G(v_h + w_h)[w_h]\big\|_{Y_h} &\le \big\|F_h(v_h + w_h) - F_h(v_h) \\ &- DF_h(v_h)[w_h]\big\|_{Y_h} + \big\|\big(DF_h(v_h) - DF(v_h + w_h)\big)[w_h]\big|_{Y_h}\end{aligned}$$

and the right-hand side converges faster to 0 than $\|w_h\|_X$ as $w_h \to 0$.
(ii) If X_h is a Hilbert space the function $F_h(v) = \|v\|_X$, $v \in X_h$, is Newton differentiable with

$$G_h(v) = \begin{cases} v/|v| & \text{if } v \neq 0, \\ \xi & \text{if } v = 0, \end{cases}$$

where $\xi \in X_h$ with $\|\xi\|_X \le 1$ is arbitrary.
(iii) The function $F_h : \mathbb{R} \to \mathbb{R}$, $s \mapsto \max\{0, s\}$, is Newton differentiable with Newton derivative $G_h(s) = 0$ for $s < 0$, $G_h(0) = \delta$ for arbitrary $\delta \in [0, 1]$, and $G(s) = 1$ for $s > 0$.
(iv) If $1 \le p < q \le \infty$, the mapping

$$F_h : L^q(\Omega) \to L^p(\Omega), \quad v \mapsto \max\{0, v(x)\}$$

is Newton differentiable with the Newton derivative $G_h(v_h)$ for G_h as above. For $p = q$ this is false.

The semismooth Newton method is similar to the classical Newton iteration but employs the Newton derivative instead of the classical derivative.

Algorithm 4.4 (*Semismooth Newton method*) *Given $u_h^0 \in X_h$, compute the sequence $(u_h^j)_{j=0,1,\dots}$ via $u_h^{j+1} = u_h^j + d_h^j$ with $d_h^j \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)^m$ such that*

$$G_h(u_h^j)[d_h^j, v_h] = -F_h(u_h^j)[v_h]$$

for all $v_h \in \mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)^m$. Stop the iteration if $\|d_h^j\|_H \le \varepsilon_{\mathrm{stop}}$ for a norm $\|\cdot\|_H$ on $\mathscr{S}^1_{\mathrm{D}}(\mathscr{T}_h)^m$.

Theorem 4.11 (Superlinear convergence) *Suppose that $F_h(u_h) = 0$ and $F_h : X_h \to Y_h$ is Newton differentiable at u_h, such that the linear mapping $G(\widetilde{u}_h) : X_h \to Y_h$ is invertible with $\|G_h^{-1}(\widetilde{u}_h)\|_{\mathrm{L}(Y_h, X_h)} \le M$ for every $\widetilde{u}_h \in B_\varepsilon(u_h)$ with some $\varepsilon > 0$. Then the semismooth Newton method converges superlinearly to u_h if u_h^0 is sufficiently close to u_h, i.e., for every $\eta > 0$, there exists $J \ge 0$ such that for all $j \ge J$, we have*

$$\|u_h^{j+1} - u_h\|_X \le \eta \|u_h^j - u_h\|_X.$$

Proof Noting $d_h^j = -G_h(u_h^j)^{-1} F_h(u_h^j)$, we have

$$u_h^{j+1} - u_h = u_h^j - G_h^{-1}(u_h^j) F_h(u_h^j) - u_h$$

$$
\begin{aligned}
&= u_h^j - u_h - G_h^{-1}(u_h^j)\big(F_h(u_h^j) - F_h(u_h)\big) \\
&= -G_h^{-1}(u_h^j)\big(F_h(u_h^j) - F_h(u_h) - G_h(u_h^j)(u_h^j - u_h)\big).
\end{aligned}
$$

Writing $u_h^{j+1} = u_h + w_h^{j+1}$, we have

$$
\begin{aligned}
\|w_h^{j+1}\|_X \le \|G_h^{-1}(u_h + w_h^j)\|_{\mathrm{L}(Y_h, X_h)} \| F_h(u_h + w_h^j) - F_h(u_h) \\
- G_h(u_h + w_h^j) w_h^j \|_{Y_h} \\
\le M\varphi(\|w_h^j\|_X)
\end{aligned}
$$

with a function $\varphi(s)$ satisfying $\varphi(s)/s \to 0$ as $s \to 0$. If $\|w_h^0\|_X$ is sufficiently small, e.g., $\|w_h^0\|_X \le \varepsilon/(M\theta)$ with $\theta = \max_{s\in[0,1]} \varphi(s)$, then we inductively find $u_h^j \in B_\varepsilon(u_h)$ for all $j \ge 0$ and $\|w_h^j\|_X \to 0$ as $j \to \infty$. For $J \ge 0$ such that $\phi(\|w_h^j\|_X) \le (\eta/M)\|w_h^j\|_X$ for all $j \ge J$, we verify the estimate of the theorem. □

Remark 4.15 If F_h is twice continuously differentiable so that $G_h = DF_h$ is locally Lipschitz continuous and $\|DF_h^{-1}(\widetilde{u}_h)\|_{\mathrm{L}(Y_h,X_h)} \le M$, then Algorithm 4.4 coincides with the classical Newton iteration which is locally and quadratically convergent.

4.3.4 Nonsmooth, Strongly Convex Minimization

For Banach spaces X and Y, proper, convex, and lower semicontinuous functionals $G : X \to \mathbb{R} \cup \{+\infty\}$, $F : Y \to \mathbb{R} \cup \{+\infty\}$, and a bounded, linear operator $\Lambda : X \to Y$, we consider the *saddle-point problem*

$$
\inf_{u\in X} \sup_{p\in Y'} \langle \Lambda u, p\rangle - F^*(p) + G(u) = \inf_{u\in X} \sup_{p\in Y'} L(u, p).
$$

The pair (u, p) is a saddle point for L if and only if

$$
\Lambda u \in \partial F^*(p), \quad -\Lambda' p \in \partial G(u),
$$

where $\Lambda' : Y' \to X'$ denotes the formal adjoint of Λ. The related primal and dual problem consist in the minimization of the functionals

$$
I(u) = F(\Lambda u) + G(u), \quad D(p) = -F^*(p) - G^*(-\Lambda' p).
$$

We have $I(u) - D(p) \ge 0$ for all $(u, p) \in X \times Y'$ with equality if and only if (u, p) is a saddle point for L. We assume in the following that X and Y are Hilbert spaces and identify them with their duals. The descent and ascent flows $\partial_t u = -\partial_u L(u, p)$ and $\partial_t p = \partial_p L(u, p)$, respectively, motivate the following algorithm. Further details about related nonsmooth minimization problems can be found in [14].

Algorithm 4.5 (*Primal-dual iteration*) *Let* $(u^0, p^0) \in X \times Y$ *and set* $d_t u^0 = 0$. *Compute the sequences* $(u^j)_{j=0,1,\dots}$ *and* $(p^j)_{j=0,1,\dots}$ *by iteratively solving the equations*

$$\begin{aligned} \widetilde{u}^{j+1} &= u^j + \tau d_t u^j, \\ &- d_t p^{j+1} + \Lambda \widetilde{u}^{j+1} \in \partial F^*(p^{j+1}), \\ &- d_t u^{j+1} - \Lambda' p^{j+1} \in \partial G(u^{j+1}). \end{aligned}$$

Stop the iteration if $\|u^{j+1} - u^j\|_X \le \varepsilon_{\text{stop}}$.

Remark 4.16 The equations in Algorithm 4.5 are equivalent to the variational inequalities

$$\begin{aligned} &\langle -d_t u^{j+1} - \Lambda' p^{j+1}, v - u^{j+1} \rangle_X \le G(v) - G(u^{j+1}) - \frac{\alpha}{2}\|v - u^{j+1}\|_X^2, \\ &\langle -d_t p^{j+1} + \Lambda \widetilde{u}^{j+1}, q - p^{j+1} \rangle_Y \le F^*(q) - F^*(p^{j+1}) \end{aligned}$$

for all $(v, q) \in X \times Y$. Here, $\alpha > 0$ if G is uniformly convex.

We prove convergence of Algorithm 4.5 assuming that $\alpha > 0$. We abbreviate by $\|\Lambda\|$ the operator norm $\|\Lambda\|_{\mathrm{L}(X,Y)}$.

Theorem 4.12 (Convergence) *Let* (u, p) *be a saddle point for* L. *If* $\tau\|\Lambda\| \le 1$, *we have for every* $J \ge 0$ *that*

$$\frac{1 - \tau\|\Lambda\|}{2}\|p - p^{J+1}\|_Y^2 + \frac{1}{2}\|u - u^{J+1}\|_X^2 + \tau \sum_{j=0}^{J} \frac{\alpha}{2}\|u - u^{j+1}\|_X^2 \le \frac{1}{2}\|p - p^0\|_Y^2 + \frac{1}{2}\|u - u^0\|_X^2.$$

In particular, the iteration of Algorithm 4.5 terminates.

Proof We denote $\delta_u^{j+1} = u - u^{j+1}$ and $\delta_p^{j+1} = p - p^{j+1}$ in the following. Using that $d_t \delta_u^{j+1} = -d_t u^{j+1}$ and $d_t \delta_p^{j+1} = -d_t p^{j+1}$, we find that

$$\begin{aligned} \Upsilon(j+1) &= \frac{d_t}{2}\big(\|\delta_p^{j+1}\|_Y^2 + \|\delta_u^{j+1}\|_X^2\big) + \frac{\tau}{2}\big(\|d_t \delta_u^{j+1}\|_X^2 + \|d_t \delta_p^{j+1}\|_Y^2\big) + \frac{\alpha}{2}\|\delta_u^{j+1}\|_X^2 \\ &= \langle d_t \delta_p^{j+1}, \delta_p^{j+1} \rangle_Y + \langle d_t \delta_u^{j+1}, \delta_u^{j+1} \rangle_X + \frac{\alpha}{2}\|\delta_u^{j+1}\|_X^2 \\ &= -\langle d_t p^{j+1}, p - p^{j+1} \rangle_Y - \langle d_t u^{j+1}, u - u^{j+1} \rangle_X + \frac{\alpha}{2}\|u - u^{j+1}\|_X^2. \end{aligned}$$

The equations for $d_t p^{j+1}$ and $d_t u^{j+1}$ of Algorithm 4.5 and their equivalent characterization in Remark 4.16 lead to

$$\begin{aligned}
\Upsilon(j+1) \le F^*(p) - F^*(p^{j+1}) - \langle \Lambda \widetilde{u}^{j+1}, p - p^{j+1}\rangle_Y \\
+ G(u) - G(u^{j+1}) + \langle \Lambda' p^{j+1}, u - u^{j+1}\rangle_X \\
= \big[\langle \Lambda u, p^{j+1}\rangle_Y - F^*(p^{j+1}) + G(u)\big] \\
- \big[\langle \Lambda u^{j+1}, p\rangle_Y - F^*(p) + G(u^{j+1})\big] + \langle \Lambda u^{j+1}, p\rangle_Y \\
- \langle \Lambda \widetilde{u}^{j+1}, p - p^{j+1}\rangle_Y - \langle \Lambda' p^{j+1}, u^{j+1}\rangle_Y.
\end{aligned}$$

The definitions of $F^{**} = F$ and G^* imply that

$$\begin{aligned}
\langle \Lambda u, p^{j+1}\rangle_Y - F^*(p^{j+1}) \le F(\Lambda u), \\
-\langle u^{j+1}, \Lambda' p\rangle_X - G(u^{j+1}) \le G^*(-\Lambda' p).
\end{aligned}$$

These estimates and the identity $u^{j+1} - \widetilde{u}^{j+1} = \tau^2 d_t^2 u^{j+1} = -\tau^2 d_t^2 \delta_u^{j+1}$ allow us to deduce that

$$\begin{aligned}
\Upsilon(j+1) \le F(\Lambda u) + G(u) + F^*(p) + G^*(-\Lambda' p) \\
+ \langle \Lambda u^{j+1}, p\rangle_Y - \langle \Lambda \widetilde{u}^{j+1}, p - p^{j+1}\rangle_Y - \langle \Lambda' p^{j+1}, u^{j+1}\rangle_X \\
= I(u) - D(p) - \tau^2 \langle \Lambda d_t^2 \delta_u^{j+1}, \delta_p^{j+1}\rangle_Y.
\end{aligned}$$

We use $I(u) - D(p) = 0$ to derive the estimate

$$\Upsilon(j+1) \le -\tau^2 \langle \Lambda d_t^2 \delta_u^{j+1}, \delta_p^{j+1}\rangle_Y.$$

A summation of the estimate over $j = 0, 1, \dots, J$ and multiplication by τ lead to

$$\begin{aligned}
\frac{1}{2}\big(\|\delta_p^{J+1}\|_Y^2 + \|\delta_u^{J+1}\|_X^2\big) + \frac{\tau^2}{2}\sum_{j=0}^{J}\big(\|d_t \delta_u^{j+1}\|_X^2 + \|d_t \delta_p^{j+1}\|_Y^2\big) + \frac{\alpha}{2}\sum_{j=0}^{J}\|\delta_u^{j+1}\|_X^2 \\
\le \frac{1}{2}\big(\|\delta_p^0\|_Y^2 + \|\delta_u^0\|_X^2\big) - \tau^3 \sum_{j=0}^{J}\langle \Lambda d_t^2 \delta_u^{j+1}, \delta_p^{j+1}\rangle.
\end{aligned}$$

A summation by parts, $-d_t \delta_u^0 = d_t u^0 = 0$, and Young's inequality show that

$$\begin{aligned}
-\tau^3 \sum_{j=0}^{J}\langle \Lambda d_t^2 \delta_u^{j+1}, \delta_p^{j+1}\rangle_Y = \tau^3 \sum_{j=0}^{J}\langle \Lambda d_t \delta_u^{j}, d_t \delta_p^{j+1}\rangle_Y + \tau^2 \langle \Lambda d_t \delta_u^j, \delta_p^j\rangle_Y\big|_{j=0}^{J+1} \\
\le \frac{\tau^2}{2}\Big(\sum_{j=0}^{J}\tau^2 \|\Lambda d_t \delta_u^j\|_Y^2 + \|d_t \delta_p^{j+1}\|_Y^2\Big) + \frac{\tau\|\Lambda\|}{2}\|\delta_p^{J+1}\|_Y^2 + \frac{\tau^3}{2\|\Lambda\|}\|\Lambda d_t \delta_u^{J+1}\|_Y^2.
\end{aligned}$$

A combination of the estimates proves the theorem. □

Remarks 4.17 (i) The assumption that a saddle point exists implies that primal and dual problem are related by a strong duality principle.
(ii) If F is strongly convex and G is only convex, then the roles of u and p have to be exchanged to ensure convergence.
(iii) The algorithm may be regarded as an inexact Uzawa algorithm. The classical Uzawa method corresponds to omitting $d_t u^{j+1}$, i.e., solving the equation $-\Lambda' p^{j+1} \in \partial G(u^{j+1})$ for u^{j+1} at every step of the algorithm.
(iv) Algorithm 4.5 is practical if the proximity operators $r = (1 + \tau \partial F^*)^{-1} q$ and $w = (1 + \tau \partial G)^{-1} v$ can be easily evaluated, i.e., if the unique minimizers of

$$w \mapsto \frac{1}{2\tau} \|w - v\|_X^2 + G(w), \quad r \mapsto \frac{1}{2\tau} \|r - q\|_Y^2 + F^*(r),$$

are directly accessible. This is the case for quadratic functionals and indicator functionals.

Example 4.22 In the case of the discretized Poisson problem with $X = \mathcal{S}_0^1(\mathcal{T}_h)$, we may choose $Y = \mathcal{L}^0(\mathcal{T}_h)^d$, $\Lambda = \nabla$,

$$F(p_h) = \frac{1}{2} \int_\Omega |p_h|^2 \,\mathrm{d}x, \quad G(u_h) = \int_\Omega f u_h \,\mathrm{d}x,$$

and exchange the roles of u_h and p_h. Letting $P_{h,0} f$ denote the L^2 projection onto $\mathcal{S}_0^1(\mathcal{T}_h)$, the iteration reads

$$\begin{aligned} \widetilde{p}_h^{j+1} &= p_h^j + \tau d_t p_h^{j+1}, \\ -d_t u_h^{j+1} + \operatorname{div}_h \widetilde{p}_h^{j+1} &= P_h f, \\ -d_t p_h^{j+1} + \nabla u_h^{j+1} &= p_h^{j+1}. \end{aligned}$$

The discrete divergence operator $\operatorname{div}_h : \mathcal{L}^0(\mathcal{T}_h)^d \to \mathcal{S}_0^1(\mathcal{T}_h)$ is for every elemen-twise constant vector field $q_h \in \mathcal{L}^0(\mathcal{T}_h)^d$ defined by $(\operatorname{div}_h q_h, v_h) = -(q_h, \nabla v_h)$ for all $v_h \in \mathcal{S}_0^1(\mathcal{T}_h)$. Convergence holds if $\tau \|\nabla\| \le 1$, where $\|\nabla\| \le ch^{-1}$.

4.3.5 Nested Iteration

The semismooth and classical Newton method can only be expected to converge if the starting value u_h^0 is sufficiently close to the discrete solution u_h. The radius of the ball around u_h which contains such starting values may depend critically on the mesh-size in the sense that it becomes smaller when the mesh is refined. Such a behavior reflects the problem that the Newton scheme may not be well-defined for the underlying continuous formulation. When a sequence of refined triangulations is used, the corresponding finite element spaces are nested, and one may use an approximate solution computed on a coarse grid to define a starting value for the iteration process on the finer grid. Besides providing a method to construct feasible starting values, this approach can also significantly reduce the computational effort.

Algorithm 4.6 (*Nested iteration*) *Let* $(\mathcal{T}_\ell)_{\ell=0,\dots,L}$ *be a sequence of triangulations with* $\mathcal{S}^1(\mathcal{T}_{\ell-1}) \subset \mathcal{S}^1(\mathcal{T}_\ell)$ *for* $\ell = 1, 2, \dots, L$. *Set* $\ell = 0$ *and choose* $u_\ell^0 \in \mathcal{S}^1(\mathcal{T}_\ell)$.
(i) Iteratively approximate a solution $u_\ell \in \mathcal{S}^1(\mathcal{T}_\ell)$ *of* $F_\ell(u_\ell) = 0$ *using the starting value* u_ℓ^0 *to obtain an approximate solution* $u_\ell^* \in \mathcal{S}^1(\mathcal{T}_\ell)$.
(ii) Stop if $\ell = L$. *Otherwise set* $u_{\ell+1}^0 = u_\ell^*$, $\ell \to \ell + 1$, *and continue with* (i).

We make the ideas more precise for a red-green-blue refinement method. The definition is easily generalized to other refinement methods such as newest-vertex bisection.

Definition 4.4 We say that $\mathcal{T}_h$ is a *refinement* of the triangulation $\mathcal{T}_H$ if $\mathcal{S}^1(\mathcal{T}_H) \subset \mathcal{S}^1(\mathcal{T}_h)$ and for every node $z^h \in \mathcal{N}_h$ we either have $z^h \in \mathcal{N}_H$ or there exist nodes $z_1^H, z_2^H \in \mathcal{N}_H$ with $z^h = (z_1^H + z_2^H)/2$, cf. Fig. 4.3.

Lemma 4.1 (Prolongation) *Let* $\mathcal{T}_h$ *be a refinement of the triangulation* $\mathcal{T}_H$. *Given* $u_H \in \mathcal{S}^1(\mathcal{T}_H)$, *we have* $u_h = u_H \in \mathcal{S}^1(\mathcal{T}_h)$ *with nodal values* $u_h(z^h) = u_H(z^h)$ *for every* $z^h \in \mathcal{N}_H \subset \mathcal{N}_h$ *and* $u_h(z^h) = (u_H(z_1^H) + u_H(z_2^H))/2$ *for every* $z^h \in \mathcal{N}_h \setminus \mathcal{N}_H$ *and* $z_1^h, z_2^h \in \mathcal{N}_H$ *with* $z^h = (z_1^H + z_2^H)/2$. *In particular, there exists a linear* prolongation operator

$$Pr^1_{H\to h} : \mathbb{R}^{\mathcal{N}_H} \to \mathbb{R}^{\mathcal{N}_h}, \quad \big(u_H(z^H)\big)_{z^H \in \mathcal{N}_H} \mapsto \big(u_H(z^h)\big)_{z^h \in \mathcal{N}_h}$$

for every $u_H \in \mathcal{S}^1(\mathcal{T}_H)$.

Proof The assertion of the lemma follows from the fact that the function u_h is affine on every one-dimensional subsimplex in the triangulation. □

Remarks 4.18 (i) The superscript 1 in $Pr^1_{H\to h}$ corresponds to affine functions. Analogously, there exists a linear operator $Pr^0_{H\to h}$ that maps the values of an elementwise constant function on $\mathcal{T}_H$ to the values of the function represented on $\mathcal{T}_h$.
(ii) Matrices that realize the linear mappings of the nodal or elementwise values are provided by the routine `red_refine.m`.
(iii) Nested iterations are the simplest version of a multigrid scheme. In more general versions, grid transfer from a fine to a coarse grid called restriction is required. This is often realized with the adjoint operators, i.e., with the transposed matrices.
(iv) For nonnested finite element spaces the grid transfer can be realized with interpolation or projection operators.

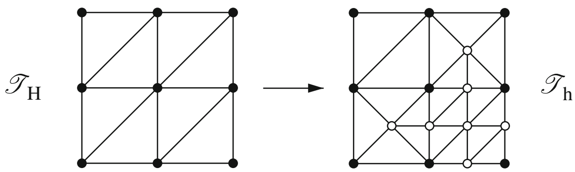

Fig. 4.3 The nodes of the refined triangulation are either existing nodes (*dots*) or midpoints of bisected edges (*circles*)

References

1. Babuška, I.: Error-bounds for finite element method. Numer. Math. **16**, 322–333 (1970/1971)
2. Berger, M.S.: Nonlinearity and Functional Analysis. Academic Press, New York (1977)
3. Boffi, D., Brezzi, F., Fortin, M.: Mixed Finite Element Methods and Applications. Springer Series in Computational Mathematics, vol. 44. Springer, Heidelberg (2013)
4. Bonnans, J.F., Gilbert, J.C., Lemaréchal, C., Sagastizábal, C.A.: Numerical Optimization, 2nd edn. Universitext. Springer, Berlin (2006)
5. Braides, A.: Approximation of Free-Discontinuity Problems. Lecture Notes in Mathematics, vol. 1694. Springer, Berlin (1998)
6. Brenner, S.C.: A Cautionary Tale in Numerical PDEs. In: Sonia Kovalevskii Lecture, ICIAM 2011. Vancouver, Canada (2011)
7. Brezzi, F.: On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers. Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge **8**(R-2), 129–151 (1974)
8. Costabel, M.: A coercive bilinear form for Maxwell's equations. J. Math. Anal. Appl. **157**(2), 527–541 (1991). http://dx.doi.org/10.1016/0022-247X(91)90104-8
9. Dacorogna, B.: Direct Methods in the Calculus of Variations. Applied Mathematical Sciences, vol. 78, 2nd edn. Springer, New York (2008)
10. Dziuk, G., Hutchinson, J.E.: Finite element approximations to surfaces of prescribed variable mean curvature. Numer. Math. **102**(4), 611–648 (2006). http://dx.doi.org/10.1007/s00211-005-0649-7
11. Evans, L.C.: Weak Convergence Methods for Nonlinear Partial Differential Equations. CBMS Regional Conference Series in Mathematics, vol. 74. Published for the Conference Board of the Mathematical Sciences, Washington (1990)
12. Ito, K., Kunisch, K.: Lagrange Multiplier Approach to Variational Problems and Applications. Advances in Design and Control, vol. 15. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (2008)
13. Nečas, J.: Sur une méthode pour résoudre les équations aux dérivées partielles du type elliptique, voisine de la variationnelle. Ann. Scuola Norm. Sup. Pisa **16**(3), 305–326 (1962)
14. Nesterov, Y., Nemirovski, A.: On first-order algorithms for ℓ_1/nuclear norm minimization. Acta Numer. **22**, 509–575 (2013)
15. Nochetto, R.H., Savaré, G., Verdi, C.: A posteriori error estimates for variable time-step discretizations of nonlinear evolution equations. Commun. Pure Appl. Math. **53**(5), 525–589 (2000)
16. Repin, S.: A Posteriori Estimates for Partial Differential Equations. Radon Series on Computational and Applied Mathematics, vol. 4. Walter de Gruyter GmbH & Co. KG, Berlin (2008)
17. Rulla, J.: Error analysis for implicit approximations to solutions to Cauchy problems. SIAM J. Numer. Anal. **33**(1), 68–87 (1996). http://dx.doi.org/10.1137/0733005
18. Struwe, M.: Variational Methods, 4th edn. Springer, Berlin (2008)