FINITE ELEMENT DISCRETIZATION OF WEIGHTED Φ-LAPLACE PROBLEMS*

ENRIQUE OTÁROLA† AND ABNER J. SALGADO‡

Abstract. We study the finite element approximation of problems involving the weighted Φ-Laplacian, where Φ is an N-function and the weight belongs to the class A_Φ. In particular, we consider a boundary value problem and an obstacle problem and derive error estimates in both cases. The analysis is based on the language of weighted Orlicz and weighted Orlicz–Sobolev spaces.

Key words. nonlinear elliptic equations, weighted Φ-Laplacian, obstacle problem, Muckenhoupt weights, finite element discretizations, quasi norms, a priori error estimates.

AMS subject classifications. 35J60, 35J70, 65N15, 65N30.

1. Introduction.

Let $d \in \mathbb{N}$ with $d \geq 2$, and let Ω be a bounded polytope in $\mathbb{R}^d$ with Lipschitz boundary $\partial\Omega$. We study finite element methods for approximating the unconstrained and constrained minimization of the following convex energy functional

$$\mathcal{J}(v) \coloneqq \int_\Omega \omega(x)\Phi(|\nabla v(x)|)\mathrm{d}x - \int_\Omega \omega(x) f(x) v(x)\mathrm{d}x, \tag{1.1}$$

over an appropriate function class that, in particular, satisfies homogeneous Dirichlet boundary conditions. Here Φ is a suitable convex function, ω is a weight belonging to a suitable class adapted to Φ, and f is a given forcing term; see Section 2 for details.

The idea behind the study of minimization problems for energies such as (1.1) is that in the resulting differential operator, i.e.,

$$D\mathcal{J}(v) = -\operatorname{div}(\boldsymbol{\mathcal{A}}(x, \nabla v)),$$

the vector field $\boldsymbol{\mathcal{A}}$ may have a nonstandard growth through Φ and is not translation invariant due to the dependence on the point x. The only simplifying assumption we make is that we consider each of these phenomena separately in the sense that

$$\boldsymbol{\mathcal{A}}(x, \boldsymbol{\zeta}) = \omega(x)\mathbf{A}(\boldsymbol{\zeta}), \qquad \mathbf{A} : \mathbb{R}^d \to \mathbb{R}^d, \qquad \mathbf{A}(\boldsymbol{\zeta}) = \Phi'(|\boldsymbol{\zeta}|)\frac{\boldsymbol{\zeta}}{|\boldsymbol{\zeta}|} \tag{1.2}$$

for a.e. $x \in \Omega$ and all $\boldsymbol{\zeta} \in \mathbb{R}^d$. The prototypical example of the type of problem we have in mind is the so-called weighted p-Laplacian: for $p \in (1, \infty)$ and $\kappa \geq 0$, we have

$$\mathbf{A}(\boldsymbol{\zeta}) = (\kappa + |\boldsymbol{\zeta}|)^{p-2}\boldsymbol{\zeta} \qquad \forall \boldsymbol{\zeta} \in \mathbb{R}^d. \tag{1.3}$$

The analysis and numerical approximation of quasilinear problems in general and the p-Laplacian in particular have a rich history, and any attempt to give a complete bibliographical account is doomed to failure. We mention here only the classical references [4] and [17]. In particular in [17] the authors introduce the language of Orlicz functions to perform an error analysis. For the approximation of obstacle problems with the p-Laplacian as operator, we can refer, for instance, to [2, 11].

*EO has been partially supported by ANID grant FONDECYT-1220156. AJS has been partially supported by NSF grant DMS-2409918.

†Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile. (enrique.otarola@usm.cl, http://eotarola.mat.utfsm.cl/)

‡Department of Mathematics, University of Tennessee, Knoxville, TN 37996, USA. (asalgad1@utk.edu, https://math.utk.edu/people/abner-j-salgado/)

On the other hand, the study of elliptic problems with nonstandard growth and inhomogeneity has received much attention in recent years, as this seems not only to adopt the difficulties of certain scenarios, but also to unify certain particular structures; see [3, 30, 13]. It is therefore only natural to investigate the approximation of solutions to such problems. This work can therefore be seen as a step towards the numerical approximation of elliptic problems in Orlicz–Musielak spaces.

Our presentation is as follows. Section 2 introduces notation and gives an overview of the main ingredients we will use. The goal is to properly describe the functional framework we use: weighted Orlicz and Orlicz–Sobolev spaces. The unconstrained minimization of (1.1) and its numerical approximation are discussed in Sections 3 and 4, respectively. The analysis hinges on the properties of the Scott-Zhang operator on weighted Orlicz spaces. Finally, an obstacle problem related to (1.1) and its numerical approximation are studied in Section 5. In this case, the analysis relies on a positivity preserving interpolant and its properties. This operator is analyzed in Section 5.3.

2. Notation and preliminary remarks. The first notation we introduce is the relation $A \lesssim B$. This means that $A \leq CB$ for a nonessential constant $C > 0$, which can change at each occurrence. The relation $A \gtrsim B$ means $B \lesssim A$, and $A \simeq B$ is the short form for $A \lesssim B \lesssim A$. If it is necessary to explicitly mention a constant C, we assume that $C > 0$ and that its value can change at each occurrence.

Let $E \subset \mathbb{R}^d$ be a measurable set, and let $|E|$ be the d-dimensional Lebesgue measure of E. If $0 < |E| < \infty$ and $v \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$, we denote the mean value of v over the set E by

$$\langle v \rangle_E := \fint_E v(x)\,\mathrm{d}x = \frac{1}{|E|}\int_E v(x)\,\mathrm{d}x.$$

Here and in the following, the relation $A := B$ indicates that A is *by definition* equal to B. The notation $B =: A$ means $A := B$.

2.1. N-functions, complementary functions, and shifted N-functions. We say that $\Phi : [0,\infty) \to [0,\infty)$ is an N-function if Φ is differentiable and its derivative Φ' satisfies the following properties: $\Phi'(0) = 0$, $\Phi'(s) > 0$ for $s > 0$, Φ' is right-continuous at any point $s \geq 0$, Φ' is nondecreasing on $[0,\infty)$, and $\lim_{s\to\infty} \Phi'(s) = \infty$ [42, Definition 4.2.1]. We note that every N-function is convex [42, Lemma 4.2.2], [34, page 7]. We say that an N-function Φ satisfies the Δ_2-condition (we shall write $\Phi \in \Delta_2$) if there exists a positive constant K such that $\Phi(2t) \leq K\Phi(t)$ for all $t \geq 0$. The smallest of these constants is denoted by $\Delta_2(\Phi)$ [42, Definition 4.4.1], [34, page 23]. Since $\Phi(t) \leq \Phi(2t)$, the Δ_2-condition guarantees that $\Phi(t) \simeq \Phi(2t)$ for all $t \geq 0$. More generally, if $\Delta_2(\Phi) < \infty$ and $a > 1$ is fixed, then $\Phi(t) \simeq \Phi(at)$ for all $t \geq 0$ [42, Exercise 4.5.5], [34, inequality (4.2)]. Finally, we mention that if Φ is an N-function and $\Delta_2(\Phi) < \infty$, then $\Phi(t) \simeq \Phi'(t)\,t$ uniformly in $t \geq 0$ [37, Proposition 7(2.6b)].

To characterize the behavior of an N-function Φ for very small and very large values of its argument, we introduce the lower and upper indices, respectively, as [33, Definition 1.1.5], [14, page 71]

$$i(\Phi) := \lim_{\lambda\downarrow 0} \frac{1}{\log\lambda} \log\left(\sup_{t>0} \frac{\Phi(\lambda t)}{\Phi(t)}\right), \qquad I(\Phi) := \lim_{\lambda\uparrow\infty} \frac{1}{\log\lambda} \log\left(\sup_{t>0} \frac{\Phi(\lambda t)}{\Phi(t)}\right). \tag{2.1}$$

It follows from the definition that $1 \leq i(\Phi) \leq I(\Phi) \leq \infty$. Moreover, if $\Phi \in \Delta_2$, then $1 \leq i(\Phi) \leq I(\Phi) < \infty$ [14, page 71]. Finally, we have that for every $\epsilon > 0$ there are constants c_ϵ and C_ϵ satisfying $0 < c_\epsilon \leq C_\epsilon$ such that, for every $s, t \geq 0$ [36, page 114],

$$c_\epsilon \min\{t^{i(\Phi)-\epsilon}, t^{I(\Phi)+\epsilon}\}\Phi(s) \leq \Phi(st) \leq C_\epsilon \max\{t^{i(\Phi)-\epsilon}, t^{I(\Phi)+\epsilon}\}\Phi(s). \tag{2.2}$$

2.1.1. Complementary functions. Let Φ be an N-function. Define $(\Phi')^{-1} : [0,\infty) \to [0,\infty)$ such that $(\Phi')^{-1}(t) := \sup\{s : \Phi'(s) \le t\}$ [42, Definition 4.3.1]. If Φ' is continuous and strictly increasing in $[0,\infty)$, then $(\Phi')^{-1}$ is the inverse function of Φ' and vice-versa [42, Remark 4.3.3]. We now define [42, Definition 4.3.1]

$$\Phi^* : [0,\infty) \to [0,\infty), \qquad \Phi^*(t) := \int_0^t (\Phi')^{-1}(s)\mathrm{d}s.$$

The function Φ^* is called the *complementary function* of Φ. We note that Φ^* is also an N-function, $(\Phi^*)'(t) = (\Phi')^{-1}(t)$ for $t > 0$, and $(\Phi^*)^* = \Phi$.

Example 1 (the p-Laplacian). Let $p \in (1,\infty)$ and $\kappa \ge 0$. We define the N-functions $\Phi_p : [0,\infty) \to [0,\infty)$ and $\Phi_{p,\kappa} : [0,\infty) \to [0,\infty)$ as

$$\Phi_p(t) := \frac{1}{p}t^p, \qquad \Phi_{p,\kappa}(t) := \int_0^t \Phi'_{p,\kappa}(s)\mathrm{d}s, \qquad \Phi'_{p,\kappa}(t) := (\kappa + t)^{p-2}t. \tag{2.3}$$

We note that both $\Phi_{p,\kappa}$ and Φ_p satisfy the Δ_2-condition with $\Delta_2(\Phi_{p,\kappa}) \le C2^{\max\{2,p\}}$ and $\Delta_2(\Phi_p) = 2^p$; see [5, page 376]. Note that the first estimate is independent of κ. The corresponding complementary functions are

$$\Phi_p^*(t) = \frac{1}{p'}t^{p'}, \qquad \Phi_{p,\kappa}^*(t) \simeq (\kappa^{p-1} + t)^{p'-2}t^2, \qquad t \in [0,\infty), \qquad \frac{1}{p} + \frac{1}{p'} = 1.$$

According to [5, page 376], we have that $\Delta_2(\Phi_p^*) = 2^{p'}$ and $\Delta_2(\Phi_{p,\kappa}^*) \le C2^{\max\{2,p'\}}$.

2.1.2. Young's inequalities. If (Φ, Φ^*) are complementary N-functions, then

$$st \le \Phi(s) + \Phi^*(t), \qquad \forall s,t \ge 0. \tag{2.4}$$

The equality holds if and only if $t = \Phi'(s)$ or $s = (\Phi')^{-1}(t)$ [42, Theorem 4.3.4]. This is commonly referred to as *Young's inequality*. We also have the following refined versions: If $\Phi, \Phi^* \in \Delta_2$, then for all $\delta > 0$ there exists $C_\delta > 0$, depending on δ, $\Delta_2(\Phi)$, and $\Delta_2(\Phi^*)$ such that [15, Lemma 32]

$$st \le \delta\Phi(s) + C_\delta\Phi^*(t) \quad \forall s,t \ge 0, \tag{2.5}$$

$$s\Phi'(t) + \Phi'(s)t \le \delta\Phi(s) + C_\delta\Phi(t) \quad \forall s,t \ge 0. \tag{2.6}$$

Note that inequality (2.4) is a generalization of the classical Young's inequality, whereas (2.5) is a generalization of "Young's inequality with δ".

2.1.3. Shifted N-functions. Given an N-function Φ with $\Phi, \Phi^* \in \Delta_2$, we introduce the family of *shifted* functions $\{\Phi_a\}_{a\ge0}$ as in [15, Definition 22]:

$$\Phi_a : [0,\infty) \to [0,\infty), \quad \Phi_a(t) := \int_0^t \Phi'_a(s)\mathrm{d}s, \quad \Phi'_a(t) := \Phi'(a+t)\frac{t}{a+t},\ t \ge 0. \tag{2.7}$$

For all $a \ge 0$, Φ_a and Φ_a^* are N-functions and $\Phi_a, \Phi_a^* \in \Delta_2$. More importantly,

$$\sup_{a\ge0}\{\Delta_2(\Phi_a)\}_{a\ge0} \cup \{\Delta_2(\Phi_a^*)\}_{a\ge0} \le C(\Delta_2(\Phi), \Delta_2(\Phi^*));$$

see [15, Lemmas 23 and 27] and [17, Lemma 6.1].

Example 2 (the shifted p-Laplacian). Recall the function $\Phi_{p,\kappa}$ introduced in (2.3). In this case we have, for $a \ge 0$ and $t \ge 0$,

$$\Phi_{p,\kappa,a}(t) \simeq (\kappa + a + t)^{p-2}t^2, \quad \Phi_{p,\kappa,a}^*(t) \simeq ((\kappa + a)^{p-1} + t)^{p'-2}t^2.$$

For all $a \ge 0$, $\Delta_2(\Phi_{p,\kappa,a}) \le C2^{\max\{2,p\}}$ and $\Delta_2(\Phi_{p,\kappa,a}^*) \le C2^{\max\{2,p\}}$. Thus, $\{\Phi_a\}_{a\ge0}$ and $\{\Phi_a^*\}_{a\ge0}$ satisfy the Δ_2-condition uniformly with respect to a; see [5, page 376].

2.2. Orlicz spaces. Let μ be a Borel measure on $\mathbb{R}^d$ that is absolutely continuous with respect to the Lebesgue measure. We denote by $L^0(\Omega)$ the set of all Lebesgue measurable functions in Ω. Let Φ be an N-function. We define $\varrho : L^0(\Omega) \to \mathbb{R}$ by

$$\varrho(v,\Phi) \coloneqq \int_\Omega \Phi(|v(x)|)\mathrm{d}\mu(x).$$

The function ϱ is a semimodular on $L^0(\Omega)$. Moreover, if Φ is positive, then ϱ is a modular [16, Lemma 2.3.10]. We define the *Orlicz space* [43, Sec. 3.1, Definition 5]

$$L^\Phi(\mu,\Omega) \coloneqq \left\{v \in L^0(\Omega) : \exists k > 0 : \varrho(kv,\Phi) < \infty\right\}$$

equipped with the Luxemburg norm (see [43, Sec. 3.2, (6) and Theorem 3])

$$\|v\|_{L^\Phi(\mu,\Omega)} \coloneqq \inf\left\{k > 0 : \varrho\left(\frac{v}{k},\Phi\right) \le 1\right\}. \tag{2.8}$$

The following is a list of properties of the space $L^\Phi(\mu,\Omega)$:

- $L^\Phi(\mu,\Omega)$ is a Banach space [43, Sec. 3.3, Theorem 10].
- If $\Phi \in \Delta_2$, then (see [43, Sec. 3.1, Theorem 2 and Sec. 3.3, Proposition 3])

$$L^\Phi(\mu,\Omega) = \left\{v \in L^0(\Omega) : \varrho(v,\Phi) < \infty\right\}.$$

- If $\Phi \in \Delta_2$, then $L^\Phi(\mu,\Omega)$ is separable [43, Sec. 3.5, Thm. 1 and Sec. 3.4, Cor. 5].
- If (Φ,Φ^*) is a pair of complementary N-functions, then we have Hölder's inequality [43, Sec. 3.3, Proposition 1 and (4)] (the constant 2 cannot be omitted):

$$\int_\Omega |v(x)w(x)|\mathrm{d}\mu(x) \le 2\|v\|_{L^\Phi(\mu,\Omega)}\|w\|_{L^{\Phi^*}(\mu,\Omega)} \qquad \forall v \in L^\Phi(\mu,\Omega),\ \forall w \in L^{\Phi^*}(\mu,\Omega).$$

- If (Φ,Φ^*) is a pair of complementary N-functions and $\Phi,\Phi^* \in \Delta_2$, then $[L^\Phi(\mu,\Omega)]^*$ is isomorphic to $L^{\Phi^*}(\mu,\Omega)$ in the following sense [43, Sec. 4.1, Theorem 7]:

$$w \in L^{\Phi^*}(\mu,\Omega) \mapsto J_w \in [L^\Phi(\mu,\Omega)]^* : \quad J_w(v) \coloneqq \int_\Omega v(x)w(x)\mathrm{d}\mu(x),\ v \in L^\Phi(\mu,\Omega),$$

and $\|J_w\|_{L^\Phi(\mu,\Omega)^*} \le 2\|w\|_{L^{\Phi^*}(\mu,\Omega)}$. $L^\Phi(\mu,\Omega)$ is reflexive [43, Sec. 4.1, Thm. 10].

For further facts and properties of Orlicz spaces and their generalizations, we refer the reader to [34, 43, 16, 29, 39].

2.3. Weights. A weight ω is a function in $L^1_{\mathrm{loc}}(\mathbb{R}^d)$ such that $\omega(x) > 0$ for a.e. $x \in \mathbb{R}^d$. For a weight ω and a measurable set $A \subset \mathbb{R}^d$, we set $\omega(A) \coloneqq \int_A \omega(x)\,\mathrm{d}x$.

Let $p \in (1,\infty)$. We say that a weight ω belongs to the Muckenhoupt class A_p if there is a positive constant C such that for every ball $B \subset \mathbb{R}^d$ [46, Definition 1.2.2]

$$\left(\fint_B \omega(x)\,\mathrm{d}x\right)\left(\fint_B \omega(x)^{-\frac{1}{p-1}}\,\mathrm{d}x\right)^{p-1} \le C. \tag{2.9}$$

The infimum over all such constants C is called the Muckenhoupt characteristic of ω and is denoted by $[\omega]_{A_p}$. We note that

$$\omega \in A_p \iff \omega' \coloneqq \omega^{-\frac{1}{p-1}} \in A_{p'}, \quad [\omega]_{A_p} = [\omega']^{p-1}_{A_{p'}}, \quad p \in (1,\infty), \quad \frac{1}{p}+\frac{1}{p'} = 1;$$

see [46, Remark 1.2.4, item 4]. Many useful properties follow from the fact that $\omega \in A_p$. We mention here two of them that will be useful for us in what follows:

- *Open ended property*: If $\omega \in A_p$ with $p \in (1,\infty)$, then there is $\delta > 0$ such that $\omega \in A_{p-\delta}$; see [46, Corollary 1.2.17] and [19, Corollary 7.6, item (2)].
- *Embedding:* If $1 < p < q < \infty$, then $A_p \subset A_q$; see [46, Remark 1.2.4, item 3].

We refer to [46, 19, 31, 27, 14] for more details on Muckenhoupt weights.

When dealing with the obstacle problem we restrict ourselves to weights that behave well near $\partial\Omega$. The following definition is inspired by [24, Definition 2.5].

DEFINITION 2.1 ($A_p(\Omega)$). *Let $\Omega \subset \mathbb{R}^d$ be bounded and Lipschitz, and let $p \in (1,\infty)$. Given $\omega \in A_p$, we say $\omega \in A_p(\Omega)$ if there is an open $\mathcal{G} \subset \Omega$, and $\varepsilon, \omega_l > 0$ such that $\{x \in \Omega : \operatorname{dist}(x,\partial\Omega) < \varepsilon\} \subset \mathcal{G}$, $\omega_{|\bar{\mathcal{G}}} \in C(\bar{\mathcal{G}})$, and $\inf_{x\in\bar{\mathcal{G}}} \omega(x) \geq \omega_l$.*

2.4. Weighted Orlicz spaces. We have finally reached the point where we can describe the functional framework that we will use in our analysis. Namely, that of *weighted Orlicz spaces*. Let ω be a weight. Define the measure μ via $\mathrm{d}\mu(x) \coloneqq \omega(x)\,\mathrm{d}x$. Given an N-function Φ, we define the weighted Orlicz space

$$L^\Phi(\omega,\Omega) \coloneqq \left\{ v \in L^0(\Omega) : \exists k > 0 :\ \varrho(kv,\Phi) \coloneqq \int_\Omega \omega(x)\Phi(k|v(x)|)\mathrm{d}x < \infty \right\}.$$

We endow $L^\Phi(\omega,\Omega)$ with the Luxemburg norm (2.8). Some properties of this space have already been described in Section 2.2. Further properties are examined below.

2.4.1. The class B_Φ. To define weak derivatives of functions in a weighted Orlicz space, such functions must be distributions. To describe when this is the case, we introduce a class of weights adapted to a given N-function Φ. We say that the weight $\omega \in B_\Phi$ if for every ball $B \subset \mathbb{R}^d$ there is a constant $\mu > 0$ such that

$$\frac{1}{\mu}\int_B \Phi^*\left(\frac{\mu}{\omega(x)}\right)\omega(x)\,\mathrm{d}x < \infty. \tag{2.10}$$

Condition (2.10) is a generalization of the condition that A. Kufner introduced in [38, (1.5)] in the context of weighted Sobolev spaces, and it occurs naturally when proving that $L^\Phi(\omega,\Omega) \subset L^1_{\mathrm{loc}}(\Omega)$. The proof of this embedding result can be found in [36, Proposition 3.6 and 3.7], where it is shown that $\omega \in B_\Phi$ is not only sufficient but also necessary for the embedding. Because of its simplicity, we give a proof of sufficiency for the sake of completeness.

PROPOSITION 2.2 ($L^\Phi(\omega,\Omega) \subset L^1_{\mathrm{loc}}(\Omega)$). *Let Φ be an N-function. If $\omega \in B_\Phi$, then every function in $L^\Phi(\omega,\Omega)$ is locally integrable in Ω.*

Proof. Let $K \subset \Omega$ be compact, let $v \in L^\Phi(\omega,\Omega)$, and let $\mu > 0$. According to Young's inequality (2.4), we have

$$\begin{aligned}\int_K |v(x)|\mathrm{d}x &= \frac{1}{\mu}\int_K \mu\omega(x)^{-1}|v(x)|\omega(x)\mathrm{d}x \\ &\leq \frac{1}{\mu}\left[\int_K \Phi^*\left(\frac{\mu}{\omega(x)}\right)\omega(x)\,\mathrm{d}x + \int_K \omega(x)\Phi(|v(x)|)\mathrm{d}x\right].\end{aligned}$$

The condition $\omega \in B_\Phi$ controls the first term on the right hand side of the previous inequality, and this shows the result. ∎

Remark 2.3 (class B_Φ). Let us consider, for $p \in (1,\infty)$, the function Φ_p introduced in (2.3). In this case we have that $\omega \in B_{\Phi_p}$ if

$$\int_B \Phi_p^*\left(\frac{\mu}{\omega(x)}\right)\omega(x)\,\mathrm{d}x = \int_B \frac{1}{p'}\left(\frac{\mu}{\omega(x)}\right)^{p'}\omega(x)\,\mathrm{d}x \simeq \int_B \omega(x)^{-\frac{1}{p-1}}\,\mathrm{d}x \leq C.$$

In the language of [38, Definition 1.4], the finiteness of the last integral for all balls $B \subset \mathbb{R}^d$ means that $\omega \in B_p$. As [38, Theorem 1.5] shows, $\omega \in B_p$ is sufficient for the inclusion $L^p(\omega,\Omega) \subset L^1_{\mathrm{loc}}(\Omega)$. In this sense, the class B_Φ stands as the natural extension of the class B_p. Note also that $A_p \subset B_p$.

2.4.2. The maximal function. Given a function $v \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$, the Hardy–Littlewood maximal function is defined as

$$M[v](x) \coloneqq \sup_{B \ni x} \fint_B |v(y)|\,\mathrm{d}y.$$

The continuity of this maximal function is a very important step in the analysis of a function space. Given an N-function Φ and a weight ω, we say that $\omega \in A_\Phi$ if and only if there is $C > 0$ such that for all $\delta > 0$ and for all balls $B \subset \mathbb{R}^d$ we have

$$\left[\fint_B \delta\omega(x)\,\mathrm{d}x\right] \Phi'\left(\fint_B (\Phi')^{-1}\left(\frac{1}{\delta\omega(x)}\right)\,\mathrm{d}x\right) \le C. \tag{2.11}$$

The significance of condition (2.11) is the content of the following result.

THEOREM 2.4 (continuity). *Let Φ be an N-function such that $\Phi, \Phi^* \in \Delta_2$, and let ω be a weight. Then, the following statements are equivalent:*

1. $\omega \in A_\Phi$.
2. $\omega \in A_{i(\Phi)}$.
3. *There is a constant $C > 0$ such that*

$$\int_{\mathbb{R}^d} \Phi(M[v](x))\omega(x)\,\mathrm{d}x \le C \int_{\mathbb{R}^d} \Phi(|v(x)|)\omega(x)\,\mathrm{d}x \quad \forall v \in L^\Phi(\omega,\mathbb{R}^d), \tag{2.12}$$

where C depends on Φ only through $\Delta_2(\{\Phi,\Phi^*\})$ and on ω only through $[\omega]_{A_{i(\Phi)}}$.

Proof. This is the content of [32, Theorem 1] and [33, Theorem 2.1.1]. The dependence on the constant C in (2.12) is not explicitly stated, but can be inferred from the proof of this result. ∎

Remark 2.5 (A_Φ is a natural generalization of the classes A_p). Let us consider the function Φ_p of (2.3) for $p \in (1,\infty)$. The weight ω belongs to A_{Φ_p} if and only if

$$\left[\fint_B \delta\omega(x)\,\mathrm{d}x\right] \Phi_p'\left(\fint_B (\Phi_p')^{-1}\left(\frac{1}{\delta\omega(x)}\right)\,\mathrm{d}x\right) = \left[\fint_B \omega(x)\,\mathrm{d}x\right]\left[\fint_B \omega(x)^{-\frac{1}{p-1}}\,\mathrm{d}x\right]^{p-1} \le C,$$

for all balls $B \subset \mathbb{R}^d$, i.e., $A_{\Phi_p} = A_p$. Thus, A_Φ is not only the natural generalization of the classes A_p, *but also the most general condition* that can be imposed to have a reasonable theory and numerical analysis for PDEs in weighted Orlicz spaces.

Remark 2.6 ($A_\Phi \subset B_\Phi$). As we have already mentioned, for any N-function Φ, $(\Phi')^{-1} = (\Phi^*)'$. Since $(\Phi^*)'$ is nondecreasing on $[0,\infty)$, $\omega \in A_\Phi$ therefore implies that

$$\begin{aligned}\frac{1}{\mu}\int_B \Phi^*\left(\frac{\mu}{\omega(x)}\right)\omega(x)\,\mathrm{d}x &= \frac{1}{\mu}\int_B \int_0^{\frac{\mu}{\omega(x)}} (\Phi^*)'(t)\,\mathrm{d}t\;\omega(x)\,\mathrm{d}x \\ &\le \frac{1}{\mu}\int_B (\Phi^*)'\left(\frac{\mu}{\omega(x)}\right)\frac{\mu}{\omega(x)}\omega(x)\,\mathrm{d}x \\ &= \int_B (\Phi^*)'\left(\frac{\mu}{\omega(x)}\right)\,\mathrm{d}x = \int_B (\Phi')^{-1}\left(\frac{\mu}{\omega(x)}\right)\,\mathrm{d}x < \infty,\end{aligned}$$

because the last integral is the argument inside Φ' in the definition of the class A_Φ. In other words, we have obtained that $A_\Phi \subset B_\Phi$; see Remark 2.3.

2.4.3. Weighed Orlicz–Sobolev spaces. Let Φ be an N-function, and let $\omega \in A_\Phi \subset B_\Phi$; see Remark 2.6. Then, as shown in Proposition 2.2, the elements of the space $L^\Phi(\omega,\Omega)$ are distributions and they have distributional derivatives. We can then define the *weighted Orlicz–Sobolev space* [39, Definition 3.1.1]

$$W^{1,\Phi}(\omega,\Omega) \coloneqq \{v \in L^\Phi(\omega,\Omega) : \partial_i v \in L^\Phi(\omega,\Omega)\ \forall i \in \{1,\dots,d\}\}$$

endowed with the norm $\|v\|_{W^{1,\Phi}(\omega,\Omega)} \coloneqq \|v\|_{L^\Phi(\omega,\Omega)} + \|\nabla v\|_{L^\Phi(\omega,\Omega)}$. Many of the properties of the classical Sobolev spaces extend to the weighted Orlicz–Sobolev setting:
1. $W^{1,\Phi}(\omega,\Omega)$ is a Banach space [29, Theorem 6.1.4(b)].
2. If $\Phi \in \Delta_2$, then $W^{1,\Phi}(\omega,\Omega)$ is separable [29, Theorem 6.1.4(c)].
3. If $\Phi, \Phi^* \in \Delta_2$, then $W^{1,\Phi}(\omega,\Omega)$ is reflexive [29, Theorem 6.1.4(d)].

We define $W_0^{1,\Phi}(\omega,\Omega)$ as the closure of $C_0^\infty(\Omega) \cap W^{1,\Phi}(\omega,\Omega)$ in $W^{1,\Phi}(\omega,\Omega)$ [29, Definition 6.1.8]. It follows that this space has the same properties as $W^{1,\Phi}(\omega,\Omega)$; see [29, Theorem 6.1.9]. In addition, the following modular Poincaré inequality holds.

PROPOSITION 2.7 (modular Poincaré inequality). *Let Φ be an N-function such that $\Phi, \Phi^* \in \Delta_2$. If $\omega \in A_\Phi$, then*

$$\int_\Omega \omega(x)\Phi(|v(x)|)\mathrm{d}x \lesssim \int_\Omega \omega(x)\Phi(|\nabla v(x)|)\mathrm{d}x \quad \forall v \in W_0^{1,\Phi}(\omega,\Omega). \tag{2.13}$$

Proof. The result follows directly from [14, Theorem 4.15]. For the sake of completeness, we provide some context and explanations.
- [14, Theorem 4.15] begins by letting $\mathcal{B}$ be a Muckenhoupt basis that is $A_{p,\mathcal{B}}$ open. In our context, $\mathcal{B}$ is nothing but the set of all balls $B \subset \mathbb{R}^d$, so that $A_{p,\mathcal{B}} = A_p$, the standard Muckenhoupt class. Thus, the fact that $A_{p,\mathcal{B}}$ is open is the open ended property of the class A_p stated at the end of Section 2.3.
- The next step is to provide a class $\mathcal{F}$ of pairs (f,g) of nonnegative measurable functions that are not identically zero, so that for some $p_0 \in [1,\infty)$ and all $w_0 \in A_{p_0}$

$$\int_{\mathbb{R}^d} f(x)^{p_0} w_0(x)\,\mathrm{d}x \le C \int_{\mathbb{R}^d} g(x)^{p_0} w_0(x)\,\mathrm{d}x.$$

  With $\mathcal{F} \coloneqq \{(|w|,|\nabla w|) : w \in C_0^\infty(\Omega)\}$, the inequality is a weighted Poincaré inequality in Ω that holds for any $p_0 \in (1,\infty)$ and all $w_0 \in A_{p_0}$ [23, Theorem 1.3].
- In [14, Theorem 4.15], it is assumed that Φ is an N-function such that $1 < i(\Phi) \le I(\Phi) < \infty$. We note that $1 < i(\Phi) \le I(\Phi) < \infty$ is equivalent to the requirement that $\Phi, \Phi^* \in \Delta_2$ [14, page 71]. Let us also note that in [14] the complementary function of Φ is denoted by $\bar{\Phi}$.
- In [14, Theorem 4.15] it is assumed that $\omega \in A_{i(\Phi)}$. Note that, according to Theorem 2.4, $\omega \in A_{i(\Phi)}$ is equivalent to the requirement that $\omega \in A_\Phi$.

We can thus conclude that (2.13) holds for all $v \in C_0^\infty(\Omega)$. Finally, we argue by density. This concludes the proof. ∎

Remark 2.8 (equivalence). On $W_0^{1,\Phi}(\omega,\Omega)$, $\|\nabla w\|_{L^\Phi(\omega,\Omega)}$ defines a norm that is equivalent to $\|w\|_{W^{1,\Phi}(\omega,\Omega)}$ provided that $\Phi, \Phi^* \in \Delta_2$ and $\omega \in A_\Phi$.

2.4.4. Weighted Lebesgue and Sobolev spaces. A specific example of the constructions described above are weighted Lebesgue and Sobolev spaces. Let $p \in (1,\infty)$, let Φ_p be as in (2.3), and let $\omega \in A_p$. We set

$$W^{1,p}(\omega,\Omega) \coloneqq W^{1,\Phi_p}(\omega,\Omega), \qquad W_0^{1,p}(\omega,\Omega) \coloneqq W_0^{1,\Phi_p}(\omega,\Omega).$$

On $W_0^{1,p}(\omega,\Omega)$ we have a weighted Poincaré inequality: if $D \subset \mathbb{R}^d$ is open, bounded, and Lipschitz, $p \in (1,\infty)$, and $\omega \in A_p$, then [23, Theorem 1.3]

$$\|w\|_{L^p(\omega,D)} \leq C_p \operatorname{diam}(D) \|\nabla w\|_{L^p(\omega,D)} \quad \forall w \in W_0^{1,p}(\omega,D). \tag{2.14}$$

The constant C_p depends on ω only through $[\omega]_{A_p}$. The following inequality is also useful for our analysis [40, Lemmas 3.1 and 4.2]: There exists $w_D \in \mathbb{R}$ such that

$$\|w - w_D\|_{L^p(\omega,D)} \leq C_p \operatorname{diam}(D) \|\nabla w\|_{L^p(\omega,D)} \quad \forall w \in W^{1,p}(\omega,D). \tag{2.15}$$

We also recall a scaled trace inequality. Let $T \subset \mathbb{R}^d$ be a simplex and $F \subset T$ a face of T. If $p \in [1,\infty)$ and $w \in W^{1,p}(T)$, then its trace $\operatorname{tr}_{\partial T} w \in L^p(\partial T)$ satisfies

$$\frac{1}{|F|} \| \operatorname{tr}_F w \|_{L^p(F)}^p \lesssim \frac{1}{|T|} \|w\|_{L^p(T)}^p + \frac{\operatorname{diam}(T)^p}{|T|} \|\nabla w\|_{L^p(T)}^p. \tag{2.16}$$

We conclude with a simple embedding result between weighted Orlicz spaces. For a more detailed discussion we refer to [42, Section 4.5] and [36, Theorem 3.8].

PROPOSITION 2.9 (embedding). *Let Φ be an N-function such that $\Phi, \Phi^* \in \Delta_2$, and let $\omega \in A_\Phi$. If $i(\Phi) > 2$, then we have the embedding $L^\Phi(\omega,\Omega) \hookrightarrow L^2(\omega,\Omega)$. Consequently, and under the same assumptions, $W_0^{1,\Phi}(\omega,\Omega) \hookrightarrow W_0^{1,2}(\omega,\Omega)$.*

Proof. Let $\varrho = \frac{1}{2}(i(\Phi) - 2) > 0$ and set $T \geq 1$. Note that, since $2 < i(\Phi) \leq I(\Phi) < \infty$, we have that $t^{i(\Phi)-\varrho} = \min\{t^{i(\Phi)-\varrho}, t^{I(\Phi)+\varrho}\}$ for $t \geq T$. In addition, since $i(\Phi) - \varrho > 2$, we also have $t^{i(\Phi)-\varrho} \geq t^2$ for all $t \geq T$. These observations yield

$$\frac{1}{2} t^2 \leq \frac{1}{2} \min\{t^{i(\Phi)-\varrho}, t^{I(\Phi)+\varrho}\} \leq \frac{1}{2\Phi(1) c_\varrho} \Phi(t) \quad \forall t \geq T, \tag{2.17}$$

where we have used (2.2). For the rest of the proof, we set $G = (2\Phi(1)c_\varrho)^{-1} > 0$. We now let $v \in L^\Phi(\omega,\Omega)$ and define $\Omega_T := \{x \in \Omega : |v(x)| \geq T\}$. We may then estimate

$$\frac{1}{2} \int_\Omega |v(x)|^2 \omega(x) \, \mathrm{d}x = \frac{1}{2} \int_{\Omega_T} |v(x)|^2 \omega(x) \, \mathrm{d}x + \frac{1}{2} \int_{\Omega \setminus \Omega_T} |v(x)|^2 \omega(x) \, \mathrm{d}x. \tag{2.18}$$

To control the first integral on the right hand side of (2.18), we use the bound we have just obtained (2.17):

$$\frac{1}{2} \int_{\Omega_T} |v(x)|^2 \omega(x) \, \mathrm{d}x \leq G \int_\Omega \Phi(|v(x)|) \omega(x) \, \mathrm{d}x.$$

To bound the second integral on the right hand side of (2.18), we use that v is "small" in $\Omega \setminus \Omega_T$, that ω is a weight, and that Ω is bounded:

$$\frac{1}{2} \int_{\Omega \setminus \Omega_T} |v(x)|^2 \omega(x) \, \mathrm{d}x \leq \frac{1}{2} T^2 \omega(\Omega).$$

We have therefore arrived at the following estimate

$$\frac{1}{2} \int_\Omega |v(x)|^2 \omega(x) \, \mathrm{d}x \lesssim \int_\Omega \Phi(|v(x)|) \omega(x) \, \mathrm{d}x + \omega(\Omega),$$

which proves the embedding. ∎

Remark 2.10 (Sobolev embeddings). In the context of the last result, it seems natural to expect that $W_0^{1,\Phi}(\omega,\Omega) \hookrightarrow L^q(\omega,\Omega)$ for some $q > i(\Phi)$. After all, when $\Phi = \Phi_p$ so that $i(\Phi) = p$, this is the content of well-known weighted Sobolev embedding results. However, to the best of our knowledge, the only reference dealing with weighted Orlicz-Sobolev embeddings is [35], where the definition of weighted Orlicz spaces is different and not equivalent to the one we adopt here.

3. The weighted Φ-Laplace problem. We now present the assumptions that allow us to properly formulate and analyze the problem of minimizing (1.1). Throughout our discussion, $\Omega \subset \mathbb{R}^d$ is a Lipschitz polytope and ω is a weight; further assumptions on ω will be made later. With respect to Φ, we assume the following; compare with [5, assumption A.1] and [6, assumption 2.1].

ASSUMPTION 3.1 (structural assumptions). *Let Φ be an N-function that satisfies the following conditions: $\Phi \in C^1([0,\infty)) \cap C^2(0,\infty)$ and*

$$\Phi'(t) \simeq t\,\Phi''(t) \tag{3.1}$$

uniformly in $t > 0$. The constants hidden in $\simeq$ are called the characteristics of Φ.

For $f \in L^{\Phi^*}(\omega,\Omega)$, the following variational problem is a necessary and sufficient condition for a minimum of $\mathcal{J}$ over $W_0^{1,\Phi}(\omega,\Omega)$: Find $u \in W_0^{1,\Phi}(\omega,\Omega)$ such that

$$\int_\Omega \omega(x)\mathbf{A}(\nabla u(x))\cdot\nabla v(x)\mathrm{d}x = \int_\Omega \omega(x) f(x) v(x)\mathrm{d}x \quad \forall v \in W_0^{1,\Phi}(\omega,\Omega), \tag{3.2}$$

where $\mathbf{A}$ is defined in (1.2). We note that, on the basis of Youngs inequality, we have

$$\left|\int_\Omega \omega(x) f(x) v(x)\mathrm{d}x\right| \le \int_\Omega \omega(x)\Phi(|v(x)|)\mathrm{d}x + \int_\Omega \omega(x)\Phi^*(|f(x)|)\mathrm{d}x < \infty,$$

so that the right hand side of (3.2) is well-defined. Assume that $\omega \in A_\Phi$. A convex minimization argument guarantees the existence and uniqueness of a weak solution to problem (3.2).

3.1. Some further properties. The following consequences of Assumption 3.1 are important. Φ is strictly convex in $(0,\infty)$, Φ' is strictly monotone increasing in $(0,\infty)$, and Φ satisfies the Δ_2-condition; see [37, Remark 42] and [6, page 487]. Moreover, $\Delta_2(\Phi)$ depends only on the characteristics of Φ. Φ^* also satisfies Assumption 3.1; see [15, Lemma 25]. In particular, $\Delta_2(\Phi^*)$ depends only on the characteristics of Φ. Uniformly in $s,t \in \mathbb{R}$, we have the following properties [15, Lemma 24]:

$$\Phi''(|s|+|t|)|s-t| \simeq \Phi'_{|s|}(|s-t|), \qquad \Phi''(|s|+|t|)|s-t|^2 \simeq \Phi_{|s|}(|s-t|). \tag{3.3}$$

The following shifted version of Young's inequality holds: for all $\delta > 0$ there exists a positive constant C_δ such that [15, Lemma 32]

$$s\Phi'_a(t) + \Phi'_a(s)t \le \delta\Phi_a(s) + C_\delta\Phi_a(t), \qquad \forall s,t,a \ge 0. \tag{3.4}$$

3.2. The operator A. Let Φ be an N-function that satisfies Assumption 3.1. We recall the definition of the vector field $\mathbf{A}$ in (1.2), namely,

$$\mathbf{A}:\mathbb{R}^d \to \mathbb{R}^d, \qquad \mathbf{A}(\boldsymbol{\zeta}) \coloneqq \Phi'(|\boldsymbol{\zeta}|)\frac{\boldsymbol{\zeta}}{|\boldsymbol{\zeta}|} \quad \forall \boldsymbol{\zeta} \in \mathbb{R}^d. \tag{3.5}$$

EXAMPLE 3 (the p-Laplacian). Let $p \in (1,\infty)$, and let $\kappa \ge 0$. Recall the function $\Phi_{p,\kappa}$ introduced in (2.3). Note that $\Phi_{p,\kappa}$ satisfies the conditions in Assumption 3.1 and that the characteristics of $\Phi_{p,\kappa}$ do not depend on κ. In fact, we have

$$\min\{1,p-1\}(\kappa+t)^{p-2} \le \Phi''_{p,\kappa}(t) \le \max\{1,p-1\}(\kappa+t)^{p-2}$$

for every $t > 0$ [5, page 376]. In this scenario

$$\mathbf{A}(\boldsymbol{\zeta}) = (\kappa+|\boldsymbol{\zeta}|)^{p-2}\boldsymbol{\zeta}, \quad \boldsymbol{\zeta} \in \mathbb{R}^d.$$

Remark 3.2 (Φ-monotonicity and Φ-growth). It is shown in [15, Lemma 21] that there exist positive constants C_0 and C_1 such that

$$(\mathbf{A}(\boldsymbol{\zeta}) - \mathbf{A}(\boldsymbol{\eta})) \cdot (\boldsymbol{\zeta} - \boldsymbol{\eta}) \geq C_0 \Phi''(|\boldsymbol{\zeta}| + |\boldsymbol{\eta}|)|\boldsymbol{\zeta} - \boldsymbol{\eta}|^2 \quad \forall \boldsymbol{\zeta}, \boldsymbol{\eta} \in \mathbb{R}^d, \tag{3.6}$$

$$|\mathbf{A}(\boldsymbol{\zeta}) - \mathbf{A}(\boldsymbol{\eta})| \leq C_1 \Phi''(|\boldsymbol{\zeta}| + |\boldsymbol{\eta}|)|\boldsymbol{\zeta} - \boldsymbol{\eta}| \quad \forall \boldsymbol{\zeta}, \boldsymbol{\eta} \in \mathbb{R}^d. \tag{3.7}$$

Here, C_0 and C_1 depend only on $\Delta_2(\Phi)$, $\Delta_2(\Phi^*)$, and the characteristics of Φ.

Let us now define the N-function $\Psi : [0, \infty) \to [0, \infty)$ by

$$\Psi(0) = 0, \qquad \Psi'(t) \coloneqq \sqrt{t\Phi'(t)}, \quad t > 0.$$

In [15, Lemma 25] it is shown that the N-functions Ψ and Ψ^* also satisfy Assumption 3.1 and that $\Psi''(t) \simeq \sqrt{\Phi''(t)}$ uniformly in $t > 0$. With the function Ψ at hand, we introduce the vector field $\mathbf{V}$ as follows:

$$\mathbf{V} : \mathbb{R}^d \to \mathbb{R}^d, \qquad \mathbf{V}(\boldsymbol{\zeta}) \coloneqq \Psi'(|\boldsymbol{\zeta}|)\frac{\boldsymbol{\zeta}}{|\boldsymbol{\zeta}|} \quad \forall \boldsymbol{\zeta} \in \mathbb{R}^d. \tag{3.8}$$

An application of [15, Lemma 21] shows that the bounds (3.6) and (3.7) hold if $\mathbf{A}$ and Φ are replaced by $\mathbf{V}$ and Ψ, respectively.

EXAMPLE 4 (the p-Laplacian). Let $p \in (1, \infty)$, and let $\kappa \geq 0$. The following functions correspond to those defined in (2.3):

$$\Psi_p'(t) = t^{\frac{p}{2}}, \quad \mathbf{V}_p(\boldsymbol{\zeta}) = |\boldsymbol{\zeta}|^{\frac{p-2}{2}}\boldsymbol{\zeta}, \quad \Psi_{p,\kappa}'(t) = (\kappa + t)^{\frac{p-2}{2}}t, \quad \mathbf{V}_{p,\kappa}(\boldsymbol{\zeta}) = (\kappa + |\boldsymbol{\zeta}|)^{\frac{p-2}{2}}\boldsymbol{\zeta}.$$

The following result expresses the relationship between $\mathbf{A}$, $\mathbf{V}$, and $\{\Phi_a\}_{a \geq 0}$.

PROPOSITION 3.3 (equivalences). *Let* $\mathbf{A}$ *and* $\mathbf{V}$ *be defined as in* (3.5) *and* (3.8)*, respectively. Then, for all* $\boldsymbol{\zeta}, \boldsymbol{\eta} \in \mathbb{R}^d$*, we have*

$$(\mathbf{A}(\boldsymbol{\zeta}) - \mathbf{A}(\boldsymbol{\eta})) \cdot (\boldsymbol{\zeta} - \boldsymbol{\eta}) \simeq |\mathbf{V}(\boldsymbol{\zeta}) - \mathbf{V}(\boldsymbol{\eta})|^2 \tag{3.9}$$

$$\simeq \Phi_{|\boldsymbol{\zeta}|}(|\boldsymbol{\zeta} - \boldsymbol{\eta}|) \tag{3.10}$$

$$\simeq \Phi_{|\boldsymbol{\eta}|}(|\boldsymbol{\zeta} - \boldsymbol{\eta}|) \tag{3.11}$$

$$\simeq |\boldsymbol{\zeta} - \boldsymbol{\eta}|^2 \Phi''(|\boldsymbol{\zeta}| + |\boldsymbol{\eta}|). \tag{3.12}$$

Moreover, for all $\boldsymbol{\zeta}, \boldsymbol{\eta} \in \mathbb{R}^d$*, we have*

$$|\mathbf{A}(\boldsymbol{\zeta}) - \mathbf{A}(\boldsymbol{\eta})| \simeq \Phi'_{|\boldsymbol{\zeta}|}(|\boldsymbol{\zeta} - \boldsymbol{\eta}|). \tag{3.13}$$

The constants hidden in $\simeq$ *only depend on the characteristics of* Φ*.*

Proof. A proof of the estimates (3.9)–(3.12) can be found in [15, Lemma 3] while a proof for (3.13) can be found in [37, Corollary 64]. ∎

4. Finite element discretization of the Φ-Laplace problem. Now that we have established the well-posedness of (3.2), we proceed with its approximation. Since we have assumed that Ω is a polytope, it can be meshed exactly. Denote by $\mathscr{T}_h = \{T\}$ a conforming partition, or mesh, of $\bar{\Omega}$ into closed simplices T of size $h_T = \operatorname{diam}(T)$. Here, $h = \max\{h_T : T \in \mathscr{T}_h\}$. For $T \in \mathscr{T}_h$, we define S_T as the set of all elements in $\mathscr{T}_h$ that share at least one vertex with T. We denote by $\mathbb{T} = \{\mathscr{T}_h\}_{h>0}$ a collection of conforming meshes that are refinements of an initial mesh $\mathscr{T}_0$. We assume that $\mathbb{T}$

satisfies the so-called shape regularity condition [22, Definition 1.107]: there exits a constant $\sigma > 1$ such that

$$\sup_{\mathscr{T}_h \in \mathbb{T}} \max\{\sigma_T : T \in \mathscr{T}_h\} \leq \sigma,$$

where $\sigma_T \coloneqq h_T/\rho_T$ is the shape coefficient of T and ρ_T is the diameter of the sphere inscribed in T. With this setting, given $\mathscr{T}_h \in \mathbb{T}$, we define the finite element spaces

$$\mathbb{W}_h \coloneqq \{v_h \in W^{1,\infty}(\Omega) : \ v_{h|T} \in \mathbb{P}_1(T) \ \forall T \in \mathscr{T}_h\}, \qquad \mathbb{V}_h \coloneqq \mathbb{W}_h \cap W_0^{1,\infty}(\Omega).$$

The finite element approximation of (3.2) is: Find $u_h \in \mathbb{V}_h$ such that

$$\int_\Omega \omega(x)\mathbf{A}(\nabla u_h(x)) \cdot \nabla v_h(x)\mathrm{d}x = \int_\Omega \omega(x) f(x) v_h(x)\mathrm{d}x \quad \forall v_h \in \mathbb{V}_h. \tag{4.1}$$

A convex minimization argument yields the existence and uniqueness of u_h.

4.1. A best approximation result. We present a best approximation result for (4.1). To derive it, we first establish the following Galerkin orthogonality property:

$$\int_\Omega \omega(x)\left(\mathbf{A}(\nabla u(x)) - \mathbf{A}(\nabla u_h(x))\right) \cdot \nabla v_h(x)\mathrm{d}x = 0 \quad \forall v_h \in \mathbb{V}_h. \tag{4.2}$$

This is the first step towards proving *best approximation* for finite element solutions.

THEOREM 4.1 (best approximation). *Let $u \in W_0^{1,\Phi}(\omega,\Omega)$ solve (3.2), and let $u_h \in \mathbb{V}_h$ be its finite element approximation defined as the solution to (4.1). Then,*

$$\|\mathbf{V}(\nabla u) - \mathbf{V}(\nabla u_h)\|_{L^2(\omega,\Omega)} \lesssim \inf_{v_h \in \mathbb{V}_h} \|\mathbf{V}(\nabla u) - \mathbf{V}(\nabla v_h)\|_{L^2(\omega,\Omega)}. \tag{4.3}$$

Proof. The derivation of the best approximation property (4.3) modifies the arguments in [17, Lemma 5.2] to account for our weighted setting. To proceed, we define $e = u - u_h$. Moreover, for an arbitrary but fixed $v_h \in \mathbb{V}_h$, we set $z = u - v_h$. We begin by using the equivalence (3.9) and Galerkin orthogonality to obtain

$$\begin{aligned}
\mathfrak{I} &\coloneqq \int_\Omega \omega(x)|\mathbf{V}(\nabla u(x)) - \mathbf{V}(\nabla u_h(x))|^2\mathrm{d}x \\
&\simeq \int_\Omega \omega(x)(\mathbf{A}(\nabla u(x)) - \mathbf{A}(\nabla u_h(x))) \cdot \nabla e(x)\mathrm{d}x \\
&= \int_\Omega \omega(x)(\mathbf{A}(\nabla u(x)) - \mathbf{A}(\nabla u_h(x))) \cdot \nabla z(x)\mathrm{d}x.
\end{aligned}$$

We now use the equivalences (3.12) and (3.3) to derive

$$\begin{aligned}
\mathfrak{I} &\simeq \int_\Omega \omega(x)\Phi''(|\nabla u(x)| + |\nabla u_h(x)|)|\nabla e(x)||\nabla z(x)|\mathrm{d}x \\
&\simeq \int_\Omega \omega(x)\Phi'_{|\nabla u(x)|}(|\nabla e(x)|)|\nabla z(x)|\mathrm{d}x.
\end{aligned}$$

An application of the shifted version of Young's inequality (3.4) shows that for all $\delta > 0$ there exists a positive constant C_δ so that

$$\begin{aligned}
\mathfrak{I} &\leq C\delta \int_\Omega \omega(x)\Phi_{|\nabla u(x)|}(|\nabla e(x)|)\mathrm{d}x + C_\delta \int_\Omega \omega(x)\Phi_{|\nabla u(x)|}(|\nabla z(x)|)\mathrm{d}x \\
&\leq C\delta \int_\Omega \omega(x)|\mathbf{V}(\nabla u(x)) - \mathbf{V}(\nabla u_h(x))|^2\mathrm{d}x + C_\delta \int_\Omega \omega(x)\Phi_{|\nabla u(x)|}(|\nabla z(x)|)\mathrm{d}x,
\end{aligned}$$

where to obtain the last inequality we used the equivalence (3.10). Using the definition of $\mathfrak{I}$ and the equivalence (3.10) once again we obtain

$$\mathfrak{I} \leq C\delta\mathfrak{I} + CC_\delta \|\mathbf{V}(\nabla u) - \mathbf{V}(\nabla v_h)\|^2_{L^2(\omega,\Omega)}.$$

The term $C\delta\mathfrak{I}$ appearing in the previous inequality can be absorbed on the left hand side if δ is chosen carefully. This gives us the estimate (4.3) and concludes the proof. ∎

4.2. Stability estimates for an interpolation operator. In this section, we derive stability estimates in weighted Orlicz–Sobolev spaces for a suitable interpolation operator Π_h. As in [17], we assume that such an interpolation operator

$$\Pi_h : W^{1,1}(\Omega) \to \mathbb{W}_h$$

satisfies the following two properties:
1. *Projection:* Π_h is a projection, i.e., $\Pi_h w_h = w_h$ for all $w_h \in \mathbb{W}_h$.
2. *Stability:* Let $v \in W^{1,1}(\Omega)$. Then, for each $T \in \mathscr{T}_h$ we have the bound

$$\|\nabla \Pi_h v\|_{L^1(T)} \lesssim \|\nabla v\|_{L^1(S_T)}. \tag{4.4}$$

Recall that S_T is the set of elements in $\mathscr{T}_h$ that share at least one vertex with T.

An example of an interpolation operator Π_h that satisfies these two properties is the so-called Scott–Zhang operator, which was introduced in [45]. The construction of Π_h is such that it preserves homogeneous boundary conditions, i.e., $v = 0$ on $\partial\Omega$ implies $\Pi_h v = 0$ on $\partial\Omega$ [45, (2.17)]. Moreover, Π_h is a projection from $W^{1,1}(\Omega)$ to $\mathbb{W}_h$ with the property that $W^{1,1}_0(\Omega)$ is mapped to $\mathbb{V}_h$ [45, Theorem 2.1]. The Scott–Zhang operator also satisfies the stability bound (4.4). We refer the reader to [45, Theorem 3.1] for a proof of this result and mention that in this case the hidden constant in (4.4) depends only on the shape-regularity coefficient σ.

Inspired by the unweighted case discussed in [17, Theorem 4.5], we present a stability result in weighted Orlicz–Sobolev spaces $W^{1,\Phi}(\omega,\Omega)$.

THEOREM 4.2 (weighted stability). *Let Φ be an N-function that satisfies Assumption 3.1, let $\omega \in A_\Phi$, and let $v \in W^{1,\Phi}(\omega,\Omega)$. Then, for $T \in \mathscr{T}_h$, we have*

$$\fint_T \omega(x)\Phi(|\nabla \Pi_h v(x)|)\mathrm{d}x \lesssim \fint_{S_T} \omega(x)\Phi(|\nabla v(x)|)\mathrm{d}x, \tag{4.5}$$

where the hidden constant depends on σ, $[\omega]_{A_{i(\Phi)}}$, and the characteristics of Φ.

Proof. Let $T \in \mathscr{T}_h$ and let $x \in T$. We begin the proof by noting that

$$|\nabla \Pi_h v(x)| \lesssim \fint_T |\nabla \Pi_h v(z)|\mathrm{d}z \leq C \fint_{S_T} |\nabla v(z)|\mathrm{d}z.$$

The first bound follows from $(\nabla \Pi_h v)_{|T} \in \mathbb{P}_0(T)$, while the second follows from the stability of the interpolation operator Π_h in (4.4). Note that C depends only on the shape-regularity coefficient σ. We can therefore rely on the monotonicity of Φ and

the Δ_2-condition, which holds because Φ satisfies Assumption 3.1, to obtain

$$\begin{aligned}\fint_T \omega(x)\Phi(|\nabla \Pi_h v(x)|)\mathrm{d}x &\leq \fint_T \omega(x)\Phi\left(C\fint_{S_T}|\nabla v(z)|\mathrm{d}z\right)\mathrm{d}x\\ &\lesssim \fint_T \omega(x)\Phi\left(\fint_{S_T}|\nabla v(z)|\mathrm{d}z\right)\mathrm{d}x\\ &= \fint_T \omega(x)\Phi\left(\fint_{S_T}|\nabla v(z)\chi_{S_T}(z)|\,\mathrm{d}z\right)\mathrm{d}x\\ &\lesssim \fint_T \omega(x)\Phi\left(M[\nabla v\chi_{S_T}](x)\right)\mathrm{d}x.\end{aligned}$$

In the last step, we used shape regularity to replace the average on the patch S_T by the average over a ball B, which is such that $x \in B$, $S_T \subset B$, and $|S_T| \approx |B|$. The average over the ball B is therefore bounded by the maximal function. Since $\omega \in A_\Phi$, we then invoke Theorem 2.4 to obtain

$$\begin{aligned}\fint_T \omega(x)\Phi(|\nabla \Pi_h v(x)|)\mathrm{d}x &\lesssim \frac{1}{|T|}\int_{\mathbb{R}^d}\omega(x)\Phi\left(M[\nabla v\chi_{S_T}](x)\right)\mathrm{d}x\\ &\lesssim \frac{1}{|T|}\int_{\mathbb{R}^d}\omega(x)\Phi\left(|\nabla v(x)\chi_{S_T}(x)|\right)\mathrm{d}x\\ &\lesssim \frac{1}{|S_T|}\int_{S_T}\omega(x)\Phi\left(|\nabla v(x)|\right)\mathrm{d}x = \fint_{S_T}\omega(x)\Phi\left(|\nabla v(x)|\right)\mathrm{d}x.\end{aligned}$$

Note that in the course of our derivations all constants depend only on σ, $[\omega]_{A_{i(\Phi)}}$, and the characteristics of Φ. This concludes the proof. ∎

We present the result of Theorem 4.2 in the context of a family of shifted functions.

COROLLARY 4.3 (weighted stability). *Let Φ be an N-function satisfying Assumption 3.1, and let $\{\Phi_a\}_{a\geq 0}$ be the family of shifted N-functions defined in (2.7) associated with Φ. If $\omega \in A_\Phi$, then we have the following bound uniformly in $a \geq 0$ and $T \in \mathscr{T}_h$:*

$$\fint_T \omega(x)\Phi_a(|\nabla \Pi_h v(x)|)\mathrm{d}x \lesssim \fint_{S_T}\omega(x)\Phi_a(|\nabla v(x)|)\mathrm{d}x, \tag{4.6}$$

where the hidden constant depends on σ, $[\omega]_{A_{i(\Phi)}}$, and the characteristics of Φ.

Proof. The desired bound results from the application of the stability bound of Theorem 4.2 to the family of shifted N-functions $\{\Phi_a\}_{a\geq 0}$ in conjunction with the fact that such a family satisfies the Δ_2-condition uniformly in $a \geq 0$ with a constant that depends only on $\Delta_2(\Phi)$; see [15, Lemma 23] and [17, Lemma 6.1]. ∎

4.3. A quasi-best approximation result for the Scott-Zhang operator. In this section, we prove that the Scott–Zhang interpolation operator Π_h satisfies the following local quasi-best approximation property. In doing so, we adapt the proof of [17, Theorem 5.7] to our weighted setting.

THEOREM 4.4 (local approximation). *Let Φ be an N-function satisfying Assumption 3.1, let $\omega \in A_\Phi$, and let $v \in W^{1,\Phi}(\omega,\Omega)$. Then, for each $T \in \mathscr{T}_h$, we have*

$$\fint_T \omega(x)|\mathbf{V}(\nabla v(x)) - \mathbf{V}(\nabla \Pi_h v(x))|^2\mathrm{d}x \lesssim \inf_{\mathbf{q}\in\mathbb{R}^d}\fint_{S_T}\omega(x)|\mathbf{V}(\nabla v(x)) - \mathbf{V}(\mathbf{q})|^2\mathrm{d}x, \tag{4.7}$$

where the hidden constant depends on σ, $[\omega]_{A_{i(\Phi)}}$, and the characteristics of Φ. In particular, if $\nabla \mathbf{V}(\nabla v)$ belongs to $L^2(\omega,\Omega)$, then we have the bound

$$\fint_T \omega(x)|\mathbf{V}(\nabla v(x)) - \mathbf{V}(\nabla \Pi_h v(x))|^2 \mathrm{d}x \lesssim h_T^2 \fint_{S_T} \omega(x)|\nabla \mathbf{V}(\nabla v)(x)|^2 \mathrm{d}x, \tag{4.8}$$

where the hidden constant depends on σ, $[\omega]_{A_{i(\Phi)}}$, and the characteristics of Φ.

Proof. We begin by noting that the equivalence (3.11) allows us to obtain

$$\int_\Omega \omega(x)|\mathbf{V}(\nabla v(x))|^2 \mathrm{d}x \leq C \int_\Omega \omega(x)\Phi(|\nabla v(x)|)\mathrm{d}x < \infty$$

because $v \in W^{1,\Phi}(\omega,\Omega)$. This implies that $\mathbf{V}(\nabla v) \in L^2(\omega,\Omega)$.

We now let $\mathbf{q} \in \mathbb{R}^d$ be arbitrary and apply the triangle inequality to obtain

$$\begin{aligned}\fint_T \omega(x)|\mathbf{V}(\nabla v(x)) - \mathbf{V}(\nabla \Pi_h v(x))|^2 \mathrm{d}x &\lesssim \fint_T \omega(x)|\mathbf{V}(\nabla v(x)) - \mathbf{V}(\mathbf{q})|^2 \mathrm{d}x \\ &\quad + \fint_T \omega(x)|\mathbf{V}(\mathbf{q}) - \mathbf{V}(\nabla \Pi_h v(x))|^2 \mathrm{d}x =: \mathbf{I}_T + \mathbf{II}_T.\end{aligned}$$

Let $\mathfrak{p} \in \mathbb{P}_1(S_T)$ be such that $\nabla \mathfrak{p} = \mathbf{q}$. Note that $\mathbf{q} = \nabla \mathfrak{p} = \nabla \Pi_h \mathfrak{p}$ because $\Pi_h \mathfrak{p} = \mathfrak{p}$. We again use the equivalence (3.10) and the previous observation to bound the term $\mathbf{II}_T$ as follows:

$$\begin{aligned}\fint_T \omega(x)|\mathbf{V}(\mathbf{q}) - \mathbf{V}(\nabla \Pi_h v(x))|^2 \mathrm{d}x &\lesssim \fint_T \omega(x)\Phi_{|\mathbf{q}|}(|\mathbf{q} - \nabla \Pi_h v(x)|)\mathrm{d}x \\ &= C \fint_T \omega(x)\Phi_{|\mathbf{q}|}(|\nabla \Pi_h \mathfrak{p}(x) - \nabla \Pi_h v(x)|)\mathrm{d}x \\ &= C \fint_T \omega(x)\Phi_{|\mathbf{q}|}(|\nabla \Pi_h (\mathfrak{p} - v)(x)|)\mathrm{d}x.\end{aligned}$$

We now apply Corollary 4.3 with $a = |\mathbf{q}|$ to the function $\mathfrak{p} - v$ and arrive at

$$\fint_T \omega(x)|\mathbf{V}(\mathbf{q}) - \mathbf{V}(\nabla \Pi_h v(x))|^2 \mathrm{d}x \lesssim \fint_{S_T} \omega(x)\Phi_{|\mathbf{q}|}(|\mathbf{q} - \nabla v(x)|)\mathrm{d}x.$$

As a final step, we apply the equivalence (3.10) once again to derive a bound for $\mathbf{II}_T$:

$$\mathbf{II}_T \lesssim \fint_{S_T} \omega(x)\Phi_{|\mathbf{q}|}(|\mathbf{q} - \nabla v(x)|)\mathrm{d}x \leq C \fint_{S_T} \omega(x)|\mathbf{V}(\mathbf{q}) - \mathbf{V}(\nabla v(x))|^2 \mathrm{d}x.$$

We now bound $\mathbf{I}_T$. To do this, we use that $T \subset S_T$ with $|T| \simeq |S_T|$ to obtain

$$\mathbf{I}_T \leq C \fint_{S_T} \omega(x)|\mathbf{V}(\mathbf{q}) - \mathbf{V}(\nabla v(x))|^2 \mathrm{d}x.$$

Combining the bounds for $\mathbf{I}_T$ and $\mathbf{II}_T$ we conclude that, for every $\mathbf{q} \in \mathbb{R}^d$,

$$\fint_T \omega(x)|\mathbf{V}(\nabla v(x)) - \mathbf{V}(\nabla \Pi_h v(x))|^2 \mathrm{d}x \leq C \fint_{S_T} \omega(x)|\mathbf{V}(\mathbf{q}) - \mathbf{V}(\nabla v(x))|^2 \mathrm{d}x. \tag{4.9}$$

This estimate immediately yields (4.7).

After we have derived the local approximation result (4.9), the a priori error estimate (4.8) follows from the weighted Poincaré inequality (2.15). Note that (4.9) holds for every $\mathbf{q} \in \mathbb{R}^d$ and that $\mathbf{V} : \mathbb{R}^d \to \mathbb{R}^d$ is surjective. This concludes the proof. ∎

4.4. An error estimate. We conclude with the following a priori error estimate, which is the main result of this section.

THEOREM 4.5 (error estimate). *Let Φ be an N-function satisfying Assumption 3.1, and let $\omega \in A_\Phi$. Let $u \in W_0^{1,\Phi}(\Omega)$ be the solution of (3.2), and let $u_h \in \mathbb{V}_h$ be its finite element approximation, which solves (4.1). If $\nabla \mathbf{V}(\nabla u) \in L^2(\omega, \Omega)$, then*

$$\|\mathbf{V}(\nabla u) - \mathbf{V}(\nabla u_h)\|_{L^2(\omega,\Omega)} \lesssim h \|\nabla \mathbf{V}(\nabla u)\|_{L^2(\omega,\Omega)}, \tag{4.10}$$

where the hidden constant depends on σ, $[\omega]_{A_{i(\Phi)}}$, and the characteristics of Φ.

Proof. The error bound (4.10) follows from the quasi-best approximation property (4.3) in conjunction with the error estimate (4.8). ∎

Remark 4.6 (regularity). Note that Theorem 4.5 assumes $\nabla \mathbf{V}(\nabla u) \in L^2(\omega, \Omega)$. In the unweighted case, i.e., $\omega \equiv 1$, such results are proved in [1, 7, 8, 15, 18].

5. The obstacle problem. We now consider the constrained minimization of $\mathcal{J}$ defined in (1.1). In the same setting and with the same assumptions as in Section 3, we additionally assume that we are given $\psi \in C^2(\bar{\Omega})$ with $\psi \le 0$ on $\partial\Omega$ and define

$$\mathcal{K}_\psi := \left\{ w \in W_0^{1,\Phi}(\omega, \Omega) : w \ge \psi \text{ a.e. in } \Omega \right\}. \tag{5.1}$$

We recall that Φ is an N-function that satisfies Assumption 3.1, and that the weight ω belongs to A_Φ. The problem we are interested in is to find $u \in \mathcal{K}_\psi$ such that

$$\mathcal{J}(u) \le \mathcal{J}(v) \quad \forall v \in \mathcal{K}_\psi.$$

A convex minimization argument [21, 26] shows that the minimizer can be equivalently characterized as the solution of the variational inequality: Find $u \in \mathcal{K}_\psi$ such that

$$\int_\Omega \omega(x) \mathbf{A}(\nabla u(x)) \cdot \nabla (u - v)(x) \, \mathrm{d}x \le \int_\Omega \omega(x) f(x) (u - v)(x) \, \mathrm{d}x \quad \forall v \in \mathcal{K}_\psi. \tag{5.2}$$

Since it will be useful for what follows, we define $\lambda \in \left(W^{1,\Phi}(\omega, \Omega)\right)'$ as

$$\langle \lambda, w \rangle = \int_\Omega \omega(x) \mathbf{A}(\nabla u(x)) \cdot \nabla w(x) \, \mathrm{d}x - \int_\Omega \omega(x) f(x) w(x) \, \mathrm{d}x. \tag{5.3}$$

We note that (5.2) can thus be rewritten as follows: Find $u \in \mathcal{K}_\psi$ such that

$$\langle \lambda, u - v \rangle \le 0 \quad \forall v \in \mathcal{K}_\psi.$$

The following result describes further properties of λ.

PROPOSITION 5.1 (properties of λ). *The functional λ defined in (5.3) defines a nonnegative Radon measure. In addition, λ also satisfies the following properties:*
- *For every $\phi \in C_0^\infty(\Omega)$ that is nonnegative, i.e., $\phi \ge 0$ in Ω, we have $\langle \lambda, \phi \rangle \ge 0$.*
- *The following* complementarity condition *holds: $\langle \lambda, u - \psi \rangle = 0$.*

Proof. Let $0 \le \phi \in C_0^\infty(\Omega)$. Define $v = u + \phi \in W_0^{1,\Phi}(\omega, \Omega)$. Since v satisfies $v \ge u \ge \psi$ in Ω, v is admissible in (5.2). We can therefore replace v in the variational inequality (5.2) and obtain $0 \ge \langle \lambda, u - (u + \phi) \rangle = -\langle \lambda, \phi \rangle$. This proves the first property. We also note that this shows that λ defines a nonnegative distribution. The Riesz-Schwartz theorem (see [44, Chapter 1, §4, Théorème V] and [47, Chapter 1, Sec. 1.7, Theorem II]) then shows that λ is a nonnegative Radon measure.

To obtain the second property, we define the coincidence and noncoincidence sets

$$\mathcal{C} = \{x \in \Omega : u(x) = \psi(x)\}, \qquad \Omega^+ = \Omega \setminus \mathcal{C} = \{x \in \Omega : u(x) > \psi(x)\}.$$

With these sets at hand, we can then write

$$\langle \lambda, u - \psi \rangle = \int_{\mathcal{C}} (u-\psi)(x)\,\mathrm{d}\lambda(x) + \int_{\Omega^+} (u-\psi)(x)\,\mathrm{d}\lambda(x) = \int_{\Omega^+} (u-\psi)(x)\,\mathrm{d}\lambda(x). \tag{5.4}$$

Now let $\mathcal{N} \subset \Omega^+$ be open, and let $\phi \in C_0^\infty(\mathcal{N})$. Consider $v = u - \epsilon\phi \in W_0^{1,\Phi}(\omega,\Omega)$. For $\epsilon > 0$, but sufficiently small, we have that $v \in \mathcal{K}_\psi$. In fact, we have that $u(x) > \psi(x)$ up to a null set in Ω^+. Then, for a sufficiently small $\epsilon > 0$, we obtain that $u(x) - \epsilon\phi(x) \geq \psi(x)$ in Ω. If we replace this function v in (5.2), we obtain

$$0 \geq \langle \lambda, u - (u - \epsilon\phi)\rangle = \epsilon\langle \lambda, \phi\rangle.$$

From this and the property that $\langle \lambda, \phi\rangle \geq 0$ for every $\phi \in C_0^\infty(\Omega)$ that is nonnegative, we conclude that $\langle \lambda, \phi\rangle = 0$ for every $\phi \in C_0^\infty(\Omega^+)$ that is nonnegative, i.e.,

$$\int_{\Omega^+} \phi(x)\,\mathrm{d}\lambda(x) = 0 \quad \forall \phi \in C_0^\infty(\Omega^+) : \ \phi \geq 0 \text{ in } \Omega^+.$$

Since $u - \psi > 0$ in Ω^+, an approximation argument and (5.4) then show that $\langle \lambda, u - \psi\rangle = 0$ as claimed. ∎

5.1. Finite element discretization. To approximate the solution of (5.2), we retain the notation and constructions from Section 4. In addition, we need to introduce the positivity preserving interpolant $\wp_h : L^1(\Omega) \to \mathbb{V}_h$, which was originally developed in [12]. Let $\mathscr{V} = \{\mathtt{v}\}$ be the collection of vertices of $\mathscr{T}_h$ and let $\mathscr{V}_{\mathrm{in}} \coloneqq \mathscr{V} \cap \Omega$. Since $\mathbb{V}_h$ consists of piecewise linears, there is a bijection between $\mathscr{V}_{\mathrm{in}}$ and the canonical Lagrange basis $\{\phi_{\mathtt{v}}\}_{\mathtt{v}\in\mathscr{V}_{\mathrm{in}}}$ of $\mathbb{V}_h$. Given $\mathtt{v} \in \mathscr{V}_{\mathrm{in}}$, we let

$$S_{\mathtt{v}} = \bigcup \{T \in \mathscr{T}_h : T \ni \mathtt{v}\}.$$

For $\mathtt{v} \in \mathscr{V}_{\mathrm{in}}$, we define $\mathcal{B}_{\mathtt{v}}$ to be the largest ball centered in $\mathtt{v}$ and contained in $S_{\mathtt{v}}$. The positivity preserving interpolant $\wp_h$ is defined as follows: for $w \in L^1(\Omega)$,

$$\wp_h w = \sum_{\mathtt{v}\in\mathscr{V}_{\mathrm{in}}} \left(\fint_{\mathcal{B}_{\mathtt{v}}} w(x)\,\mathrm{d}x \right) \phi_{\mathtt{v}}. \tag{5.5}$$

Clearly, $w \geq 0$ in Ω implies $\wp_h w \geq 0$ in Ω. Further properties will be detailed below.

With the interpolant $\wp_h$ at hand, we can define the discrete admissible set

$$\mathcal{K}_{h,\psi} \coloneqq \{w_h \in \mathbb{V}_h : w_h \geq \wp_h\psi \text{ in } \Omega\}. \tag{5.6}$$

The finite element approximation of the solution to (5.2) is $u_h \in \mathcal{K}_{h,\psi}$, which satisfies the following variational inequality for every $v_h \in \mathcal{K}_{h,\psi}$:

$$\int_\Omega \omega(x)\mathbf{A}(\nabla u_h(x)) \cdot \nabla(u_h - v_h)(x)\,\mathrm{d}x \leq \int_\Omega \omega(x) f(x)(u_h - v_h)(x)\,\mathrm{d}x. \tag{5.7}$$

Standard arguments can be used to prove the existence and uniqueness of a solution of (5.7).

5.2. A priori error bounds. We now analyze error bounds for our scheme.

PROPOSITION 5.2 (first error bound). *Let $u \in \mathcal{K}_\psi$ and $u_h \in \mathcal{K}_{h,\psi}$ be the solutions of (5.2) and (5.7), respectively. If $i(\Phi) > 2$ and $\lambda \in L^2(\omega^{-1}, \Omega)$, then we have*

$$\|\mathbf{V}(\nabla u) - \mathbf{V}(\nabla u_h)\|^2_{L^2(\omega,\Omega)} \lesssim \|\lambda\|_{L^2(\omega^{-1},\Omega)} \| [u-\psi] - \wp_h [u-\psi] \|_{L^2(\omega,\Omega)} + \int_\Omega \omega(x) |\mathbf{V}(\nabla u(x)) - \mathbf{V}(\nabla \wp_h u(x))|^2 \,\mathrm{d}x,$$

where $\wp_h$ is the positivity preserving interpolant defined in (5.5). In the previous estimate, the hidden constant is independent of h.

Proof. We first note that since $i(\Phi) > 2$, Proposition 2.9 guarantees that $u \in W^{1,2}_0(\omega, \Omega)$. Next, we set some notation. We define the error $e := u - u_h$ and the auxiliary function $z := u - \wp_h u$. We note that $\wp_h u \in \mathcal{K}_{h,\psi}$.

With the help of (3.9) we can now derive the following bound:

$$\|\mathbf{V}(\nabla u) - \mathbf{V}(\nabla u_h)\|^2_{L^2(\omega,\Omega)} \lesssim \int_\Omega \omega(x) \left[\mathbf{A}(\nabla u(x)) - \mathbf{A}(\nabla u_h(x))\right] \cdot \nabla e(x) \,\mathrm{d}x =: \mathfrak{D}.$$

We further decompose $\mathfrak{D}$ as follows:

$$\begin{aligned}\mathfrak{D} &= \int_\Omega \omega(x) \left[\mathbf{A}(\nabla u(x)) - \mathbf{A}(\nabla u_h(x))\right] \cdot \nabla z(x) \,\mathrm{d}x \\ &\quad + \int_\Omega \omega(x) \left[\mathbf{A}(\nabla u(x)) - \mathbf{A}(\nabla u_h(x))\right] \cdot \nabla (\wp_h u - u_h)(x) \,\mathrm{d}x =: \mathfrak{L} + \mathfrak{N}.\end{aligned}$$

Let us first bound the term $\mathfrak{N}$. To do this, we use the discrete variational inequality (5.7) and obtain

$$\begin{aligned}\mathfrak{N} &= \int_\Omega \omega(x) \mathbf{A}(\nabla u(x)) \cdot \nabla (\wp_h u - u_h)(x) \,\mathrm{d}x \\ &\quad + \int_\Omega \omega(x) \mathbf{A}(\nabla u_h(x)) \cdot \nabla (u_h - \wp_h u)(x) \,\mathrm{d}x \\ &\leq \int_\Omega \omega(x) \mathbf{A}(\nabla u(x)) \cdot \nabla (\wp_h u - u_h)(x) \,\mathrm{d}x + \int_\Omega \omega(x) f(x) (u_h - \wp_h u)(x) \,\mathrm{d}x \\ &= \langle \lambda, \wp_h u - u_h \rangle,\end{aligned}$$

where λ is defined in (5.3). We can then continue and write

$$\begin{aligned}\mathfrak{N} &\leq \langle \lambda, u - \psi \rangle + \langle \lambda, \wp_h \psi - u_h \rangle + \langle \lambda, [\wp_h u - \wp_h \psi] - [u - \psi] \rangle \\ &= \langle \lambda, \wp_h \psi - u_h \rangle + \langle \lambda, [\wp_h u - \wp_h \psi] - [u - \psi] \rangle ,\end{aligned}$$

where we have used the complementarity condition $\langle \lambda, u - \psi \rangle = 0$ from Proposition 5.1. Let us now examine each of the remaining terms separately.

- Since $u_h \in \mathcal{K}_{h,\psi}$, we have $u_h \geq \wp_h \psi$. This and the fact that λ is a nonnegative Radon measure (see Proposition 5.1) allows us to obtain

$$\langle \lambda, \wp_h \psi - u_h \rangle = \int_\Omega (\wp_h \psi - u_h)(x) \,\mathrm{d}\lambda(x) \leq 0.$$

- Since we have assumed that $\lambda \in L^2(\omega^{-1}, \Omega)$, the fact that $\wp_h$ is linear shows that

$$\begin{aligned}\langle \lambda, [\wp_h u - \wp_h \psi] - [u - \psi] \rangle &= \langle \lambda, \wp_h [u - \psi] - [u - \psi] \rangle \\ &\leq \|\lambda\|_{L^2(\omega^{-1},\Omega)} \| [u-\psi] - \wp_h [u-\psi] \|_{L^2(\omega,\Omega)}.\end{aligned}$$

To obtain the last estimate, we used Hölder's inequality.

To summarize, we have come to the following conclusion:

$$\mathfrak{N} \leq \|\lambda\|_{L^2(\omega^{-1},\Omega)} \| \left[u-\psi\right] - \wp_h \left[u-\psi\right] \|_{L^2(\omega,\Omega)}.$$

To estimate the term $\mathfrak{L}$, we can repeat the arguments we used to control the term $\mathfrak{I}$ in the proof of Theorem 4.1. These arguments lead us to the conclusion that

$$\mathfrak{L} \leq C\delta \|\mathbf{V}(\nabla u) - \mathbf{V}(\nabla u_h)\|^2_{L^2(\omega,\Omega)} + C_\delta \int_\Omega \omega(x) \Phi_{|\nabla u(x)|}(|\nabla z(x)|)\,\mathrm{d}x.$$

If we insert the estimates for $\mathfrak{L}$ and $\mathfrak{N}$ into the bound in which $\mathfrak{D}$ is defined, we obtain

$$\begin{aligned}
&\|\mathbf{V}(\nabla u) - \mathbf{V}(\nabla u_h)\|^2_{L^2(\omega,\Omega)} \leq C\delta \|\mathbf{V}(\nabla u) - \mathbf{V}(\nabla u_h)\|^2_{L^2(\omega,\Omega)} \\
&+ C_\delta \int_\Omega \omega(x) \Phi_{|\nabla u(x)|}(|\nabla z(x)|)\,\mathrm{d}x + C\|\lambda\|_{L^2(\omega^{-1},\Omega)} \| \left[u-\psi\right] - \wp_h \left[u-\psi\right] \|_{L^2(\omega,\Omega)}.
\end{aligned}$$

The term $\|\mathbf{V}(\nabla u) - \mathbf{V}(\nabla u_h)\|^2_{L^2(\omega,\Omega)}$ appearing on the right hand side of the previous bound can be absorbed on the left hand side if δ is chosen carefully. This shows that

$$\begin{aligned}
\|\mathbf{V}(\nabla u) - \mathbf{V}(\nabla u_h)\|^2_{L^2(\omega,\Omega)} \lesssim \|\lambda\|_{L^2(\omega^{-1},\Omega)} \| \left[u-\psi\right] - \wp_h \left[u-\psi\right] \|_{L^2(\omega,\Omega)} \\
+ \int_\Omega \omega(x) \Phi_{|\nabla u(x)|}(|\nabla z(x)|)\,\mathrm{d}x, \qquad z = u - \wp_h u.
\end{aligned}$$

Use the equivalences (3.9)–(3.10) one last time to obtain the assertion and conclude. ∎

5.3. The positivity preserving interpolant. Note that up to this point no property other than positivity has been required of $\wp_h$. In the following, however, the stability and approximation properties of $\wp_h$ are required. Ideally, we would like to have an operator that is a projection and satisfies (4.4). Note, however, that $\wp_h$ is not a projection. In fact, it is not possible to construct a positivity preserving projection; see [41]. For this reason, we need to prove suitable properties for $\wp_h$.

LEMMA 5.3 (properties of $\wp_h$). *The interpolant $\wp_h$ defined in (5.5) satisfies the following properties.*

1. *$\wp_h$ is positivity preserving: If $w \geq 0$ in Ω, then $\wp_h w \geq 0$ in Ω.*
2. *$\wp_h$ is symmetric in the sense that: If $\mathtt{v} \in \mathscr{V}_{\mathrm{in}}$ and $w_{|\mathcal{B}_\mathtt{v}} \in \mathbb{P}_1$, then $\wp_h w(\mathtt{v}) = w(\mathtt{v})$.*
3. *$\wp_h$ is locally invariant: For every $T \in \mathscr{T}_h$ with $T \cap \partial\Omega = \emptyset$, if $w_{|S_T} \in \mathbb{P}_1$, then $\wp_h w_{|T} = w_{|T}$.*
4. *$\wp_h$ is locally stable: For all $p \in [1,\infty]$ and every $T \in \mathscr{T}_h$, we have*

$$\begin{aligned}
\|\wp_h w\|_{L^p(T)} &\lesssim \|w\|_{L^p(S_T)} \quad \forall w \in L^p(\Omega), \\
\|\nabla \wp_h w\|_{L^p(T)} &\lesssim \|\nabla w\|_{L^p(S_T)} \quad \forall w \in W^{1,p}_0(\Omega).
\end{aligned}$$

5. *$\wp_h$ is weighted Orlicz stable: Let Φ be an N-function that satisfies Assumption 3.1, and let $\omega \in A_\Phi$. Then, for every $T \in \mathscr{T}_h$ and for all $a \geq 0$,*

$$⨍_T \omega(x) \Phi_a(|\nabla \wp_h w(x)|)\,\mathrm{d}x \lesssim ⨍_{S_T} \omega(x) \Phi_a(|\nabla w(x)|)\,\mathrm{d}x,$$

   where the hidden constant does not depend on T or a.
6. *$\wp_h$ has weighted approximation properties: If $\omega \in A_2$ and $w \in W^{1,2}_0(\omega,\Omega)$, then, for all $T \in \mathscr{T}_h$,*

$$\|w - \wp_h w\|_{L^2(\omega,T)} \lesssim h_T \|\nabla w\|_{L^2(\omega,S_T)}.$$

Proof. We examine each statement individually.

1. This property follows directly from the definition of $\wp_h$.
2. Since $w_{|\mathcal{B}_\mathrm{v}} \in \mathbb{P}_1$, we write $w(x) = a + \mathbf{b}\cdot(x-\mathrm{v})$ for $a \in \mathbb{R}$ and $\mathbf{b} \in \mathbb{R}^d$. We have
$$\wp_h w(\mathrm{v}) = \fint_{\mathcal{B}_\mathrm{v}} [a + \mathbf{b}\cdot(x-\mathrm{v})]\,\mathrm{d}x = a + \mathbf{b}\cdot\fint_{\mathcal{B}_\mathrm{v}} (x-\mathrm{v})\,\mathrm{d}x = a = w(\mathrm{v}).$$
3. This follows from the repetition of the previous calculation for each vertex of T, which forms an unisolvent set for $\mathbb{P}_1(T)$.
4. This is proved in [12, Lemma 3.1].
5. This follows from the preceding stability properties by repeating the arguments in the proofs of Theorem 4.2 and Corollary 4.3.
6. Let $T \in \mathscr{T}_h$, and let $w \in W_0^{1,2}(\omega,\Omega)$. Let $\mathfrak{p} \in \mathbb{P}_1(\mathbb{R}^d)$ be arbitrary. We begin the proof by noting that if we can show the bound
$$\|\wp_h w - \mathfrak{p}\|_{L^2(\omega,T)} \lesssim \|w-\mathfrak{p}\|_{L^2(\omega,S_T)} + h_T\|\nabla(w-\mathfrak{p})\|_{L^2(\omega,S_T)}, \tag{5.8}$$
then a simple application of the triangle inequality shows
$$\|w - \wp_h w\|_{L^2(\omega,T)} \lesssim \|w-\mathfrak{p}\|_{L^2(\omega,S_T)} + h_T\|\nabla(w-\mathfrak{p})\|_{L^2(\omega,S_T)}.$$
A suitable choice of $\mathfrak{p}$ shall then imply the thesis.
To prove (5.8), we proceed differently depending on whether T is an interior element or T touches the boundary. If T is an interior element, i.e., $T \cap \partial\Omega = \emptyset$, then local invariance and stability in $L^2(\omega,T)$ imply the result. On the other hand, if the element T touches the boundary we proceed as follows. Let $\{\mathrm{v}_i^T\}_{i=0}^d$ be the collection of vertices of T. Since $(\wp_h w - \mathfrak{p})_{|T} \in \mathbb{P}_1$, we can write
$$(\wp_h w - \mathfrak{p})_{|T} = \sum_{i=0}^d (\wp_h w - \mathfrak{p})(\mathrm{v}_i^T)\phi_{\mathrm{v}_i^T}.$$
We now use that $\|\phi_\mathrm{v}\|_{L^\infty(\Omega)} = 1$ for all $\mathrm{v} \in \mathscr{V}$ to obtain the estimate
$$\|\wp_h w - \mathfrak{p}\|^2_{L^2(\omega,T)} \lesssim \max_{i=0}^d \left|(\wp_h w - \mathfrak{p})(\mathrm{v}_i^T)\right|^2 \omega(T).$$
Now we distinguish two cases. If $\mathrm{v}_i^T \in \mathscr{V}_\mathrm{in}$, then
$$\begin{aligned}\left|(\wp_h w - \mathfrak{p})(\mathrm{v}_i^T)\right| = \left|\wp_h(w-\mathfrak{p})(\mathrm{v}_i^T)\right| &\le \fint_{\mathcal{B}_{\mathrm{v}_i^T}} |w(x)-\mathfrak{p}(x)|\,\mathrm{d}x \\ &\lesssim h_T^{-d}\omega^{-1}(T)^{\frac{1}{2}}\|w-\mathfrak{p}\|_{L^2(\omega,S_T)}.\end{aligned}$$
On the other hand, if $\mathrm{v}_i^T \in \partial\Omega$, then $\wp_h w(\mathrm{v}_i^T) = 0$ by definition. Let $F \subset T$ be the $(d-1)$-dimensional subsimplex containing v_i^T such that $F \subset \partial\Omega$. We thus have
$$\left|(\wp_h w - \mathfrak{p})(\mathrm{v}_i^T)\right| = |\mathfrak{p}(\mathrm{v}_i^T)| \lesssim \fint_F |\mathfrak{p}(x)|\,\mathrm{d}x = \fint_F |w(x)-\mathfrak{p}(x)|\,\mathrm{d}x,$$
where we have used that $w_{|F} = 0$. The scaled trace inequality (2.16) then yields
$$\begin{aligned}\fint_F |w(x)-\mathfrak{p}(x)|\,\mathrm{d}x &\lesssim \fint_{S_T} |w(x)-\mathfrak{p}(x)|\,\mathrm{d}x + h_T\fint_{S_T}|\nabla(w-\mathfrak{p})(x)|\,\mathrm{d}x \\ &\le h_T^{-d}\omega^{-1}(S_T)^{\frac{1}{2}}\left[\|w-\mathfrak{p}\|_{L^2(\omega,S_T)} + h_T\|\nabla(w-\mathfrak{p})\|_{L^2(\omega,S_T)}\right].\end{aligned}$$

To summarize, we have arrived at the following estimate:

$$\|\wp_h w - \mathfrak{p}\|^2_{L^2(\omega,T)} \lesssim \frac{\omega(T)\omega^{-1}(S_T)}{h_T^{2d}} \left[\|w-\mathfrak{p}\|_{L^2(\omega,S_T)} + h_T\|\nabla(w-\mathfrak{p})\|_{L^2(\omega,S_T)}\right]^2$$
$$\lesssim [\omega]_{A_2}\left[\|w-\mathfrak{p}\|_{L^2(\omega,S_T)} + h_T\|\nabla(w-\mathfrak{p})\|_{L^2(\omega,S_T)}\right]^2.$$

All properties have been proved. This concludes the proof. ∎

Next, we obtain an analogue of the bound (4.7).

PROPOSITION 5.4 (interpolation error estimate). *Let Φ be an N-function that satisfies Assumption 3.1, and let $\omega \in A_{i(\Phi)}(\Omega)$. Let us assume that $\mathscr{T}_h = \{T\}$ is such that no simplex T has more than one $(d-1)$-dimensional subsimplex on $\partial\Omega$. If $w \in W^{1,\Phi}_0(\omega,\Omega)$ is such that $\nabla\mathbf{V}(\nabla w) \in L^2(\omega,\Omega)$ and h is sufficiently small, then*

$$\|\mathbf{V}(\nabla w) - \mathbf{V}(\nabla \wp_h w)\|_{L^2(\omega,\Omega)} \lesssim h\|\nabla\mathbf{V}(\nabla w)\|_{L^2(\omega,\Omega)},$$

where the constant depends on $[\omega]_{A_{i(\Phi)}}$ and $\|\omega\|_{L^\infty(\mathcal{G})}$, but is independent of h.

Proof. We begin by noting that for every $T \in \mathscr{T}_h$ and every $\mathfrak{p} \in \mathbb{P}_1(\mathbb{R}^d)$, we have

$$\int_T \omega(x)|\mathbf{V}(\nabla w(x)) - \mathbf{V}(\nabla\wp_h w)|^2\,\mathrm{d}x \lesssim \int_T \omega(x)|\mathbf{V}(\nabla w(x)) - \mathbf{V}(\nabla\mathfrak{p})|^2\,\mathrm{d}x$$
$$+ \int_T \omega(x)|\mathbf{V}(\nabla\mathfrak{p}) - \mathbf{V}(\nabla\wp_h w)|^2\,\mathrm{d}x =: \mathbf{I} + \mathbf{II}.$$

Define the sets $\mathscr{T}_h^{\mathrm{in}} := \{T \in \mathscr{T}_h : T \cap \partial\Omega = \emptyset\}$ and $\mathscr{T}_h^{\partial} := \{T \in \mathscr{T}_h : T \cap \partial\Omega \neq \emptyset\}$. We now partition $\mathscr{T}_h = \mathscr{T}_h^{\mathrm{in}} \sqcup \mathscr{T}_h^{\partial}$.

Let $T \in \mathscr{T}_h^{\mathrm{in}}$. The local invariance of $\wp_h$, proved in Lemma 5.3, implies that $\wp_h\mathfrak{p}_{|T} = \mathfrak{p}_{|T}$. Using the equivalences (3.9)–(3.10) and the weighted Orlicz stability of $\wp_h$, which was derived in Lemma 5.3, we obtain the estimate

$$\mathbf{II} \lesssim \int_T \omega(x)\Phi_{|\nabla\mathfrak{p}|}(|\nabla\wp_h(w-\mathfrak{p})(x)|)\,\mathrm{d}x \lesssim \int_{S_T}\omega(x)\Phi_{|\nabla\mathfrak{p}|}(|\nabla(w-\mathfrak{p})(x)|\,\mathrm{d}x$$
$$\lesssim \int_{S_T}\omega(x)|\mathbf{V}(\nabla w(x)) - \mathbf{V}(\nabla\mathfrak{p})|^2\,\mathrm{d}x.$$

It follows that for every $T \in \mathscr{T}_h^{\mathrm{in}}$ the following bound applies:

$$\int_T \omega(x)|\mathbf{V}(\nabla w(x)) - \mathbf{V}(\nabla\wp_h w)|^2\,\mathrm{d}x \lesssim \int_{S_T}\omega(x)|\mathbf{V}(\nabla w(x)) - \mathbf{V}(\nabla\mathfrak{p})|^2\,\mathrm{d}x,$$

where $\mathfrak{p} \in \mathbb{P}_1(\mathbb{R}^d)$ is arbitrary. Let $\mathbf{Q} \in \mathbb{R}^d$ and choose $\mathbf{z} = \nabla\mathfrak{p} \in \mathbb{R}^d$, so that $\mathbf{V}(\mathbf{z}) = \mathbf{Q}$. This is possible because $\mathbf{V}$ is surjective. This gives us for every $T \in \mathscr{T}_h^{\mathrm{in}}$

$$\int_T \omega(x)|\mathbf{V}(\nabla w(x)) - \mathbf{V}(\nabla\wp_h w)|^2\,\mathrm{d}x \lesssim \int_{S_T}\omega(x)|\mathbf{V}(\nabla w(x)) - \mathbf{Q}|^2\,\mathrm{d}x$$
$$\lesssim h_T^2\int_{S_T}\omega(x)|\nabla\mathbf{V}(\nabla w(x))|^2\,\mathrm{d}x, \tag{5.9}$$

where we used the weighted Poincaré inequality (2.15) in the last step, by choosing the vector $\mathbf{Q}$ accordingly, and that by shape regularity $\mathrm{diam}(T) \simeq \mathrm{diam}(S_T)$.

We now consider $T \in \mathscr{T}_h^\partial$. If h is sufficiently small, it is not a mistake to assume that $T \subset \mathcal{G}$. Recall that $\mathcal{G}$ is as in Definition 2.1. Let $F \subset T$ be the unique $(d-1)$-dimensional subsimplex such that $F = T \cap \partial\Omega$ and denote by $\boldsymbol{\zeta}$ the unit outer normal to T on F. Now let $\{\mathtt{v}_i^T\}_{i=0}^d$ be the set of vertices of T and assume that $\{\mathtt{v}_i^T\}_{i=0}^d$ is numbered such that $F = \operatorname{conv}\{\mathtt{v}_i^T\}_{i=0}^{d-1}$. Define $\mathfrak{p} \in \mathbb{P}_1(\mathbb{R}^d)$ as $\mathfrak{p}(x) = b\boldsymbol{\zeta} \cdot (x - \mathtt{v}_0^T)$, where $b \in \mathbb{R}$ is to be chosen. We note that, by construction, $\mathfrak{p}_{|F} = 0$. Moreover, we have by definition that $\wp_h \mathfrak{p}_{|F} = 0$. Finally, the symmetry of $\wp_h$ proved in Lemma 5.3 implies that $\wp_h \mathfrak{p}(\mathtt{v}_d^T) = \mathfrak{p}(\mathtt{v}_d^T)$. From this follows that $\wp_h \mathfrak{p}_{|T} = \mathfrak{p}_{|T}$.

Based on the constructions described above, the weighted Orlicz stability of $\wp_h$ (see Lemma 5.3) allows us to treat **II** as before and obtain that, for every $T \in \mathscr{T}_h^\partial$,

$$\int_T \omega(x)|\mathbf{V}(\nabla w(x)) - \mathbf{V}(\nabla \wp_h w)|^2 \,\mathrm{d}x \lesssim \int_{S_T} \omega(x)|\mathbf{V}(\nabla w(x)) - \mathbf{V}(\nabla \mathfrak{p})|^2 \,\mathrm{d}x,$$

where $\mathfrak{p} \in \mathbb{P}_1(\mathbb{R}^d)$ is such that $\mathfrak{p}_{|F} = 0$, but is otherwise arbitrary.

Let us now use the fact that $w_{|F} = 0$ to deduce that ∇w and $\boldsymbol{\zeta}$ are parallel on F. This implies that $\mathbf{V}(\nabla w(x))_{|F} = v(x)\boldsymbol{\zeta}$ for some $v : F \to \mathbb{R}$. Now, given $x \in T$, we denote by $\hat{x} \in F$ the projection of x onto F. We note that, due to shape regularity, $|\mathtt{v}_d^T - \hat{\mathtt{v}}_d^T| \simeq h_T$. Define Q to be the prism with base F and height $H := |\mathtt{v}_d^T - \hat{\mathtt{v}}_d^T|$. We may now extend v to Q via

$$v(x) = v(\hat{x}) \quad \forall x \in Q.$$

As a final preparatory step, we note that $\mathbf{V}(\nabla \mathfrak{p}) = \kappa\boldsymbol{\zeta}$, where $\kappa \in \mathbb{R}$ can be chosen arbitrarily by suitably specifying the value $\mathfrak{p}(\mathtt{v}_d^T)$. We now estimate

$$\mathbf{I} \lesssim \int_T \omega(x)|\mathbf{V}(\nabla w(x)) - v(x)\boldsymbol{\zeta}|^2 \,\mathrm{d}x + \int_T \omega(x)|v(x) - \kappa|^2 \,\mathrm{d}x =: \mathbf{I}_1 + \mathbf{I}_2.$$

In order to bound $\mathbf{I}_1$, we set some notation. We let $\mathbf{w}(x) = \mathbf{V}(\nabla w(x))$. Then,

$$\begin{aligned}\mathbf{I}_1 &\le \int_Q \omega(x)|\mathbf{w}(x) - \mathbf{w}(\hat{x})|^2 \,\mathrm{d}x = \int_Q \omega(x) \left| \int_0^1 \nabla \mathbf{w}(tx + (1-t)\hat{x}) \cdot (x - \hat{x}) \,\mathrm{d}t \right|^2 \,\mathrm{d}x \\ &\lesssim \|\omega\|_{L^\infty(\mathcal{G})} \int_Q \int_0^1 \left|\nabla \mathbf{w}(tx + (1-t)\hat{x}) \cdot (x - \hat{x})\right|^2 \,\mathrm{d}t \,\mathrm{d}x,\end{aligned}$$

where we have used the fact that ω belongs to the reduced class $A_{i(\Phi)}(\Omega)$ (see Definition 2.1). We now introduce a coordinate system in which the first axis is aligned with $\boldsymbol{\zeta}$ and Q is on the half space defined by $\{x \in \mathbb{R}^d : x_1 \ge 0\}$. We thus have that, if $x \in Q$, then $x = (x_1, x')^\top$ with $x_1 \ge 0$ and $x' \in F \subset \mathbb{R}^{d-1}$. We also note that $\hat{x} = (0, x')^\top$, $x - \hat{x} = (x_1, 0)^\top$, and $y = tx + (1-t)\hat{x} = (tx_1, x')^\top$ (see Figure 5.1). With these observations in mind, we continue the bound for the term $\mathbf{I}_1$ as follows:

$$\begin{aligned}\mathbf{I}_1 &\lesssim \|\omega\|_{L^\infty(\mathcal{G})} \int_Q \int_0^1 |\partial_1 \mathbf{w}(tx_1, x')|^2 |x_1|^2 \,\mathrm{d}t \,\mathrm{d}x \\ &= \|\omega\|_{L^\infty(\mathcal{G})} \int_0^H x_1^2 \int_F \int_0^1 |\partial_1 \mathbf{w}(tx_1, x')|^2 \,\mathrm{d}t \,\mathrm{d}x' \,\mathrm{d}x_1 \\ &= \|\omega\|_{L^\infty(\mathcal{G})} \int_0^H x_1 \int_F \int_0^{x_1} |\partial_1 \mathbf{w}(r, x')|^2 \,\mathrm{d}r \,\mathrm{d}x' \,\mathrm{d}x_1 \\ &\le \|\omega\|_{L^\infty(\mathcal{G})} \int_0^H x_1 \int_F \int_0^H |\partial_1 \mathbf{w}(r, x')|^2 \,\mathrm{d}r \,\mathrm{d}x' \,\mathrm{d}x_1,\end{aligned}$$

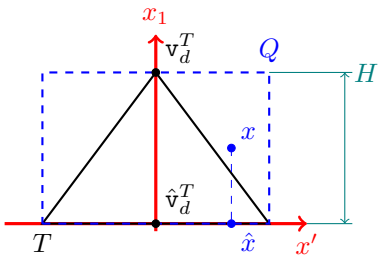

FIG. 5.1. *The geometric setting for the bound for* $\mathbf{I}_1$ *in the proof of Proposition 5.4. The element* $T \in \mathscr{T}_h^\partial$ *is shown in* ***solid black****, while the related prism is in* ***dashed blue****. We also show a generic point* $x \in Q$ *and its projection onto* F*, i.e.,* $\hat{x} \in F$*.*

where we have introduced the change of variables $r = tx_1$ in the innermost integral and used the fact that if $x \in Q$, then $x_1 \in [0, H]$. Now we just integrate with respect to x_1 and use that $H \simeq h_T$ to obtain

$$\begin{aligned}
\mathbf{I}_1 &\lesssim \frac{1}{2} h_T^2 \|\omega\|_{L^\infty(\mathcal{G})} \int_F \int_0^H |\partial_1 \mathbf{w}(r, x')|^2 \,\mathrm{d}r \,\mathrm{d}x' \\
&\leq \frac{1}{2} h_T^2 \|\omega\|_{L^\infty(\mathcal{G})} \int_F \int_0^H |\nabla \mathbf{w}(r, x')|^2 \,\mathrm{d}r \,\mathrm{d}x' \\
&= \frac{1}{2} h_T^2 \|\omega\|_{L^\infty(\mathcal{G})} \int_Q |\nabla \mathbf{w}(y)|^2 \,\mathrm{d}y \\
&\leq \frac{1}{2} h_T^2 \|\omega\|_{L^\infty(\mathcal{G})} \|\omega^{-1}\|_{L^\infty(Q)} \int_Q \omega(y) |\nabla \mathbf{V}(\nabla w(y))|^2 \,\mathrm{d}y.
\end{aligned}$$

We recall that $Q = [0, H] \times F$ and $\mathbf{w}(y) = \mathbf{V}(\nabla \mathbf{w}(y))$.

To bound $\mathbf{I}_2$, we recall that $H = |\mathrm{v}_d^T - \hat{\mathrm{v}}_d^T|$ and obtain that

$$\begin{aligned}
\mathbf{I}_2 &\leq \|\omega\|_{L^\infty(\mathcal{G})} \int_0^H \int_F |v(\hat{x}) - \kappa|^2 \,\mathrm{d}\hat{x} \,\mathrm{d}t \lesssim h_T \|\omega\|_{L^\infty(\mathcal{G})} \int_F |\operatorname{tr}_F \mathbf{w}(\hat{x}) - \kappa \boldsymbol{\zeta}|^2 \,\mathrm{d}\hat{x} \\
&\lesssim h_T \|\omega\|_{L^\infty(\mathcal{G})} \left[h_T^{-1} \int_T |\mathbf{w}(x) - \kappa \boldsymbol{\zeta}|^2 \,\mathrm{d}x + h_T \int_T |\nabla(\mathbf{w}(x) - \kappa \boldsymbol{\zeta})|^2 \,\mathrm{d}x \right] \\
&\lesssim \|\omega\|_{L^\infty(\mathcal{G})} \|\omega^{-1}\|_{L^\infty(T)} \left[\int_T \omega(x) |\mathbf{w}(x) - \kappa \boldsymbol{\zeta}|^2 \,\mathrm{d}x + h_T^2 \int_T \omega(x) |\nabla \mathbf{w}(x)|^2 \,\mathrm{d}x \right],
\end{aligned}$$

where we have used the scaled trace inequality (2.16). The weighted Poincaré inequality then shows that for every $T \in \mathscr{T}_h^\partial$,

$$\int_T \omega(x) |\mathbf{V}(\nabla w(x)) - \mathbf{V}(\nabla \wp_h w)|^2 \,\mathrm{d}x \lesssim h_T^2 \int_{S_T} \omega(x) |\nabla \mathbf{V}(\nabla w(x))|^2 \,\mathrm{d}x, \tag{5.10}$$

where the hidden constant depends on $[\omega]_{A_{i(\Phi)}}$ and $\|\omega\|_{L^\infty(\mathcal{G})}$.

It remains to add (5.9) and (5.10) and use shape regularity to obtain the claim. ∎

5.4. An error estimate. We conclude with the following error estimate for the obstacle problem. Recall that we have assumed that the obstacle satisfies $\psi \in C^2(\bar{\Omega})$.

THEOREM 5.5 (error estimate). *Let* Φ *be an* N*-function with* $i(\Phi) > 2$ *that satisfies Assumption 3.1, and let* $\omega \in A_2(\Omega)$*. Let* $u \in W_0^{1,\Phi}(\omega, \Omega)$ *and* $u_h \in \mathbb{V}_h$ *solve*

(5.2) and (5.7), respectively. Assume that $\mathbf{V}(\nabla u) \in W^{1,2}(\omega,\Omega)$*, and that* λ*, defined in (5.3), belongs to* $L^2(\omega^{-1},\Omega)$*. If* $\mathscr{T}_h$ *is such that no simplex has more than one* $(d-1)$*-dimensional subsimplex on* $\partial\Omega$ *and* h *is sufficiently small, then*

$$\|\mathbf{V}(\nabla u)-\mathbf{V}(\nabla u_h)\|_{L^2(\omega,\Omega)} \lesssim h\|\nabla\mathbf{V}(\nabla u)\|_{L^2(\omega,\Omega)} + h^{\frac{1}{2}}\|\nabla(u-\psi)\|_{L^2(\omega,\Omega)},$$

where the hidden constant is independent of h*.*

Proof. We begin with the observation that, since $i(\Phi)>2$, we have that $A_2 \subset A_{i(\Phi)}$, and thus that $\omega \in A_{i(\Phi)}(\Omega)$. We may thus apply Proposition 5.4 to the conclusion of Proposition 5.2 and obtain

$$\|\mathbf{V}(\nabla u)-\mathbf{V}(\nabla u_h)\|^2_{L^2(\omega,\Omega)} \lesssim h^2\|\nabla\mathbf{V}(\nabla u)\|^2_{L^2(\omega,\Omega)} + \|[u-\psi]-\wp_h[u-\psi]\|_{L^2(\omega,\Omega)}.$$

It remains to use the approximation property of the positivity preserving interpolant $\wp_h$ proved in Lemma 5.3 to be able to conclude. ∎

Remark 5.6 (regularity). Note that the error estimate depends on the assumption that $\mathbf{V}(\nabla u) \in W^{1,2}(\omega,\Omega)$. In addition, we need the regularity assumption $\lambda \in L^2(\omega^{-1},\Omega)$. We leave the exploration of these properties as open conjectures and refer the reader to [9, 10, 20, 25, 28, 48] for some results in this direction.

REFERENCES

[1] A. K. Balci, L. Diening, and M. Weimar, *Higher order Calderón-Zygmund estimates for the p-Laplace equation*, J. Differential Equations, 268 (2020), pp. 590–635, https://doi.org/10.1016/j.jde.2019.08.009.

[2] L. Banz, B. P. Lamichhane, and E. P. Stephan, *Higher order FEM for the obstacle problem of the p-Laplacian—a variational inequality approach*, Comput. Math. Appl., 76 (2018), pp. 1639–1660, https://doi.org/10.1016/j.camwa.2018.07.016.

[3] G. Barletta, *Existence and regularity results for nonlinear elliptic equations in Orlicz spaces*, NoDEA Nonlinear Differential Equations Appl., 31 (2024), pp. Paper No. 29, 36, https://doi.org/10.1007/s00030-024-00922-x.

[4] J. W. Barrett and W. B. Liu, *Finite element approximation of the p-Laplacian*, Math. Comp., 61 (1993), pp. 523–537, https://doi.org/10.2307/2153239.

[5] L. Belenki, L. C. Berselli, L. Diening, and M. Růžička, *On the finite element approximation of p-Stokes systems*, SIAM J. Numer. Anal., 50 (2012), pp. 373–397, https://doi.org/10.1137/10080436X.

[6] L. Belenki, L. Diening, and C. Kreuzer, *Optimality of an adaptive finite element method for the p-Laplacian equation*, IMA J. Numer. Anal., 32 (2012), pp. 484–510, https://doi.org/10.1093/imanum/drr016.

[7] S.-S. Byun and H.-S. Lee, *Optimal regularity for elliptic equations with measurable nonlinearities under nonstandard growth*, Int. Math. Res. Not. IMRN, (2024), pp. 423–461, https://doi.org/10.1093/imrn/rnad040.

[8] S.-S. Byun and M. Lim, *Global gradient estimates of very weak solutions for a general class of quasilinear elliptic equations*, J. Geom. Anal., 33 (2023), pp. Paper No. 156, 32, https://doi.org/10.1007/s12220-023-01210-3.

[9] S.-S. Byun and C. Namkyeong, *Higher differentiability for solutions of a general class of nonlinear elliptic obstacle problems with Orlicz growth*, NoDEA Nonlinear Differential Equations Appl., 29 (2022), pp. Paper No. 73, 25, https://doi.org/10.1007/s00030-022-00807-x.

[10] S.-S. Byun, K. Song, and Y. Youn, *Fractional differentiability for elliptic double obstacle problems with measure data*, Z. Anal. Anwend., 42 (2023), pp. 37–64, https://doi.org/10.4171/zaa/1721.

[11] R. Capitanelli and M. A. Vivaldi, *FEM for quasilinear obstacle problems in bad domains*, ESAIM Math. Model. Numer. Anal., 51 (2017), pp. 2465–2485, https://doi.org/10.1051/m2an/2017033.

[12] Z. Chen and R. H. Nochetto, *Residual type a posteriori error estimates for elliptic obstacle problems*, Numer. Math., 84 (2000), pp. 527–548, https://doi.org/10.1007/s002110050009.

[13] I. Chlebicka, *A pocket guide to nonlinear differential equations in Musielak-Orlicz spaces*, Nonlinear Anal., 175 (2018), pp. 1–27, https://doi.org/10.1016/j.na.2018.05.003.

[14] D. V. Cruz-Uribe, J. M. Martell, and C. Pérez, *Weights, extrapolation and the theory of Rubio de Francia*, vol. 215 of Operator Theory: Advances and Applications, Birkhäuser/Springer Basel AG, Basel, 2011, https://doi.org/10.1007/978-3-0348-0072-3.

[15] L. Diening and F. Ettwein, *Fractional estimates for non-differentiable elliptic systems with general growth*, Forum Math., 20 (2008), pp. 523–556, https://doi.org/10.1515/FORUM.2008.027.

[16] L. Diening, P. Harjulehto, P. Hästö, and M. Ružička, *Lebesgue and Sobolev spaces with variable exponents*, vol. 2017 of Lecture Notes in Mathematics, Springer, Heidelberg, 2011, https://doi.org/10.1007/978-3-642-18363-8.

[17] L. Diening and M. Růžička, *Interpolation operators in Orlicz-Sobolev spaces*, Numer. Math., 107 (2007), pp. 107–129, https://doi.org/10.1007/s00211-007-0079-9.

[18] L. Diening, B. Stroffolini, and A. Verde, *Everywhere regularity of functionals with ϕ-growth*, Manuscripta Math., 129 (2009), pp. 449–481, https://doi.org/10.1007/s00229-009-0277-0.

[19] J. Duoandikoetxea, *Fourier analysis*, vol. 29 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2001. Translated and revised from the 1995 Spanish original by David Cruz-Uribe.

[20] M. Eleuteri and A. Passarelli di Napoli, *Higher differentiability for solutions to a class of obstacle problems*, Calc. Var. Partial Differential Equations, 57 (2018), pp. Paper No. 115, 29, https://doi.org/10.1007/s00526-018-1387-x.

[21] M. Eleuteri and A. Passarelli di Napoli, *On the validity of variational inequalities for obstacle problems with non-standard growth*, Ann. Fenn. Math., 47 (2022), pp. 395–416, https://doi.org/10.54330/afm.114655.

[22] A. Ern and J.-L. Guermond, *Theory and practice of finite elements*, vol. 159 of Applied Mathematical Sciences, Springer-Verlag, New York, 2004, https://doi.org/10.1007/978-1-4757-4355-5.

[23] E. B. Fabes, C. E. Kenig, and R. P. Serapioni, *The local regularity of solutions of degenerate elliptic equations*, Comm. Partial Differential Equations, 7 (1982), pp. 77–116, https://doi.org/10.1080/03605308208820218.

[24] R. Farwig and H. Sohr, *Weighted L^q-theory for the Stokes resolvent in exterior domains*, J. Math. Soc. Japan, 49 (1997), pp. 251–288, https://doi.org/10.2969/jmsj/04920251.

[25] A. Gentile, R. Giova, and A. Torricelli, *Regularity results for bounded solutions to obstacle problems with non-standard growth conditions*, Mediterr. J. Math., 19 (2022), pp. Paper No. 270, 29, https://doi.org/10.1007/s00009-022-02162-8.

[26] J.-P. Gossez and V. Mustonen, *Variational inequalities in Orlicz-Sobolev spaces*, Nonlinear Anal., 11 (1987), pp. 379–392, https://doi.org/10.1016/0362-546X(87)90053-8.

[27] L. Grafakos, *Classical Fourier analysis*, vol. 249 of Graduate Texts in Mathematics, Springer, New York, third ed., 2014, https://doi.org/10.1007/978-1-4939-1194-3.

[28] A. G. Grimaldi and E. Ipocoana, *Higher fractional differentiability for solutions to a class of obstacle problems with non-standard growth conditions*, Adv. Calc. Var., 16 (2023), pp. 935–960, https://doi.org/10.1515/acv-2021-0074.

[29] P. Harjulehto and P. Hästö, *Orlicz spaces and generalized Orlicz spaces*, vol. 2236 of Lecture Notes in Mathematics, Springer, Cham, 2019, https://doi.org/10.1007/978-3-030-15100-3.

[30] P. Harjulehto, P. Hästö, and R. Klén, *Generalized Orlicz spaces and related PDE*, Nonlinear Anal., 143 (2016), pp. 155–173, https://doi.org/10.1016/j.na.2016.05.002.

[31] J. Heinonen, T. Kilpeläinen, and O. Martio, *Nonlinear potential theory of degenerate elliptic equations*, Dover Publications, Inc., Mineola, NY, 2006. Unabridged republication of the 1993 original.

[32] R. A. Kerman and A. Torchinsky, *Integral inequalities with weights for the Hardy maximal function*, Studia Math., 71 (1981/82), pp. 277–284, https://doi.org/10.4064/sm-71-3-277-284.

[33] V. Kokilashvili and M. Krbec, *Weighted inequalities in Lorentz and Orlicz spaces*, World Scientific Publishing Co., Inc., River Edge, NJ, 1991, https://doi.org/10.1142/9789814360302.

[34] M. A. Krasnoselskiĭ and J. B. Rutickiĭ, *Convex functions and Orlicz spaces*, P. Noordhoff Ltd., Groningen, russian ed., 1961.

[35] M. Krbec, B. Opic, and L. s. Pick, *Imbedding theorems for weighted Orlicz-Sobolev spaces*, J. London Math. Soc. (2), 46 (1992), pp. 543–556, https://doi.org/10.1112/jlms/s2-46.3.543.

[36] M. Krbec and L. Pick, *On imbeddings between weighted Orlicz spaces*, Z. Anal. Anwendungen, 10 (1991), pp. 107–117, https://doi.org/10.4171/ZAA/435.

[37] C. Kreuzer, *A convergent adaptive Uzawa finite element method for the nonlinear Stokes problem*, PhD thesis, Universität Augsburg, 2008, https://www.ruhr-uni-bochum.de/imperia/

md/content/mathematik/Kreuzer/kreuzer-dissertation.pdf. Thesis (Ph.D.)–Universitat Augsburg.

[38] A. Kufner and B. Opic, *How to define reasonably weighted Sobolev spaces*, Comment. Math. Univ. Carolin., 25 (1984), pp. 537–554.

[39] O. Méndez and J. Lang, *Analysis on function spaces of Musielak-Orlicz type*, Monographs and Research Notes in Mathematics, CRC Press, Boca Raton, FL, 2019, https://doi.org/10.1201/9781498762618.

[40] R. H. Nochetto, E. Otárola, and A. J. Salgado, *Piecewise polynomial interpolation in Muckenhoupt weighted Sobolev spaces and applications*, Numer. Math., 132 (2016), pp. 85–130, https://doi.org/10.1007/s00211-015-0709-6.

[41] R. H. Nochetto and L. B. Wahlbin, *Positivity preserving finite element approximation*, Math. Comp., 71 (2002), pp. 1405–1419, https://doi.org/10.1090/S0025-5718-01-01369-2.

[42] L. Pick, A. Kufner, O. John, and S. Fučík, *Function spaces. Vol. 1*, vol. 14 of De Gruyter Series in Nonlinear Analysis and Applications, Walter de Gruyter & Co., Berlin, extended ed., 2013.

[43] M. M. Rao and Z. D. Ren, *Theory of Orlicz spaces*, vol. 146 of Monographs and Textbooks in Pure and Applied Mathematics, Marcel Dekker, Inc., New York, 1991.

[44] L. Schwartz, *Théorie des distributions*, vol. IX-X of Publications de l'Institut de Mathématique de l'Université de Strasbourg, Hermann, Paris, 1966. Nouvelle édition, entiérement corrigée, refondue et augmentée.

[45] L. R. Scott and S. Zhang, *Finite element interpolation of nonsmooth functions satisfying boundary conditions*, Math. Comp., 54 (1990), pp. 483–493, https://doi.org/10.2307/2008497.

[46] B. O. Turesson, *Nonlinear potential theory and weighted Sobolev spaces*, vol. 1736 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2000, https://doi.org/10.1007/BFb0103908.

[47] V. S. Vladimirov, *Methods of the theory of generalized functions*, vol. 6 of Analytical Methods and Special Functions, Taylor & Francis, London, 2002.

[48] Q. Xiong, Z. Zhang, and L. Ma, *Gradient potential estimates in elliptic obstacle problems with Orlicz growth*, Calc. Var. Partial Differential Equations, 61 (2022), pp. Paper No. 83, 33, https://doi.org/10.1007/s00526-022-02196-6.