Weierstraß-Institut

für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

A semismooth Newton method with analytical path-following for the H^1-projection onto the Gibbs simplex

Lukáš Adam[1], Michael Hintermüller[1,2], Thomas M. Surowiec[3]

submitted: November 28, 2016

[1] Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin
Germany
E-Mail: adam@utia.cas.cz
hint@math.hu-berlin.de

[2] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: michael.hintermueller@wias-berlin.de

[3] Philipps-Universität Marburg
FB12 Mathematik und Informatik
Hans-Meerwein Straße 6, Lahnberge
35032 Marburg
Germany
E-Mail: surowiec@mathematik.uni-marburg.de

No. 2340

Berlin 2016

2010 *Mathematics Subject Classification.* 49M15, 90C20.

Key words and phrases. Gibbs simplex, metric projection, semismooth Newton, path-following, Ginzburg-Landau energy, multiphase field models, inpainting, data classification.

This work was carried out in the framework of the DFG under grant no. HI 1466/7-1 “Free Boundary Problems and Level Set Methods” as well as the Research Center MATHEON supported by the Einstein Foundation Berlin within projects OT1, SE5 and SE15.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

An efficient, function-space-based second-order method for the H^1-projection onto the Gibbs-simplex is presented. The method makes use of the theory of semismooth Newton methods in function spaces as well as Moreau-Yosida regularization and techniques from parametric optimization. A path-following technique is considered for the regularization parameter updates. A rigorous first and second-order sensitivity analysis of the value function for the regularized problem is provided to justify the update scheme. The viability of the algorithm is then demonstrated for two applications found in the literature: binary image inpainting and labeled data classification. In both cases, the algorithm exhibits mesh-independent behavior.

1 Introduction

In many modern applications such as image inpainting [2, 9], topology optimization [4, 5], multi-class data segmentation [12], and multiphase fluid dynamics [24, 22], the associated variational problems contain multiple phase-field functions, denoted here by $u_1, \dots, u_n$, that fulfill the conditions $u_i \geq 0$ and $\sum u_i = 1$ (in a pointwise almost everywhere (a.e.) sense). This typically leads to situations in which one needs to project onto the so-called Gibbs-simplex in the Sobolev space $H^1(\Omega)$. Mathematically speaking, this involves the solution of the following infinite dimensional optimization problem:

$$\min \left\{ \frac{1}{2}\|u - \varphi\|_{\mathbb{V}}^2 \text{ over } u \in \mathbb{V} \;\middle|\; u \in \mathcal{G} \right\}, \tag{P}$$

where $\mathcal{G} \subset H^1(\Omega)$, given by

$$\mathcal{G} := \left\{ u \in \mathbb{V} \;\middle|\; \mathbf{1}^\top u = 1,\ u_i \geq 0, i = 1, \dots, n \right\},$$

is the Gibbs simplex. Here, we let $\Omega \subset \mathbb{R}^n$ be non-empty, open, and bounded with Lipschitz boundary Γ, and set $V := H^1(\Omega)$ and $\mathbb{V} := V^n$. In addition, we fix some $\varphi \in \mathbb{V}$ and note that $\mathbf{1}$ is the column vector of ones in $\mathbb{R}^n$. The conditions on the functions u in $\mathcal{G}$ are always understood pointwise a.e. Moreover, since Ω is bounded, $\mathcal{G}$ is in fact a non-empty, closed, and convex subset in $H^1(\Omega)^n \cap L^\infty(\Omega)^n$, which is also bounded in $L^\infty(\Omega)^n$.

Although the constraint set $\mathcal{G}$ is formulated with pointwise a.e.-constraints, it is crucial that the correct norm, in this case the H^1-norm, is used in the projection. Indeed, let Γ be smooth, $n = 1$, and $\varphi \in H^2(\Omega)$ such that $\nabla\varphi \cdot n|_\Gamma = 0$. Then (P) reduces to solving the obstacle problem in H^1 given by

$$\min \left\{ \frac{1}{2}\|u\|_{\mathbb{V}}^2 - (f, u)_{L^2} \text{ over } u \in \mathbb{V} \;\middle|\; 0 \leq u \leq 1 \right\},$$

where $f = -\Delta\varphi + \varphi \in L^2(\Omega)$. It is well-known, cf. [8, 18], that the unique optimal solution $\overline{u}$ is in $H^2(\Omega)$. Furthermore, the pointwise/thresholding/L^2-projection of φ, given by

$$\varphi + \max(0, -\varphi) - \max(0, \varphi - 1),$$

is feasible, but only in $H^1(\Omega)$; as the pointwise $\max(0, \cdot)$-operator does not preserve H^2-regularity. Hence, the pointwise projection of φ onto the Gibbs simplex is feasible, yet it cannot be the solution to the original problem.

Following the approach in [16, 15], we approximate (P) by replacing $\mathcal{G}$ with a Moreau-Yosida-type regularization of the associated indicator function. This yields the approximation:

$$\theta(\gamma) := \min \left\{ \frac{1}{2}\|u-\varphi\|_{\mathbb{V}}^2 + \frac{\gamma}{2}\sum_{i=1}^{n}\int_\Omega (-u_i)_+^2\,dx + \frac{\gamma}{2}\int_\Omega \left|\mathbf{1}^\top u - 1\right|^2 dx \text{ over } u \in \mathbb{V} \right\} \quad (\mathsf{P}_\gamma)$$

The associated first-order system can be solved by a semismooth Newton method in function space. In order to increase numerical efficiency by avoiding excess parameter updates, we derive a full differential sensitivity analysis of the optimal value function $\theta(\gamma)$. These results lead to an explicit parameter update scheme referred to as "path-following." We note that our analysis is much more compact than [16, 15]. Moreover, no assumptions on the active sets are needed to derive second-order properties of $\theta(\gamma)$.

A further important factor, which motivates this work, arises from the fact that a first-discretize-then-optimize approach will typically exhibit mesh dependence under adaptive refinements (cf. the examples in Section 5). The main culprit for this aspect is due to a lack of multiplier regularity for the constraints in $\mathcal{G}$. By regularizing the problem, we increase its regularity, which enables us to obtain superlinear function-space convergence for the semismooth Newton methods for the regularized problems. Moreover, by deriving an efficient approximation of the primal-dual path value functional θ, we avoid problems connected with ill-conditioning.

The rest of the paper is structured as follows. In Section 2 we analyze the parameter dependent problem (P_γ), which lays the foundation for the path-following. Afterwards, we suggest a path-following scheme and present the algorithm in Sections 3 and 4. Finally, in Section 5, we demonstrate the performance of the algorithm on two applications taken from the literature. This is done for both path-following schemes and a comparison is made to a mesh-dependent scheme based on the direct solution of the first-order system of (P).

2 Sensitivity Analysis

As indicated in the introduction, we need to analyze the first and second-order differentiability properties of the parameter dependent quantities related to (P_γ) in order to derive a path-following scheme. We thus begin this section with some basic facts about the optimal solution mapping and optimal value functions associated with (P) and (P_γ). Throughout the discussion, we use $\bullet$ to indicate the dependence on the penalty parameter $\gamma > 0$.

Theorem 2.1. *Under the standing assumptions, the following facts hold:*

1 (P) *has a unique solution* $\bar{u}$ *and, for all* $\gamma > 0$, (P$_\gamma$) *has a unique solution* $\bar{u}^\gamma$.

2 *The optimal solution mapping* $\bar{u}^\bullet : (0,\infty) \to \mathbb{V}$ *is globally Hölder continuous with exponent* $1/2$ *and locally Lipschitz continuous.*

3 *For any sequence* $\gamma_k \to +\infty$, *the sequence* $\{\bar{u}^k\}$ *with* $\bar{u}^k := \bar{u}^{\gamma_k}$ *converges strongly to* $\bar{u}$ *in* $\mathbb{V}$.

4 *The optimal value function* $\theta : (0,\infty) \to \mathbb{R}$ *is non-negative, monotonically increasing with* $\lim_{\gamma\downarrow 0}\theta(\gamma) = 0$, *locally Lipschitz continuous, and bounded from above by* θ_∞, *the optimal value of* (P).

Proof. To see that 1. holds, consider that both (P) and (P$_\gamma$) involve the minimization of a strongly convex continuous functional over a non-empty closed set in a real Hilbert space, both problems admit unique solutions $\bar{u}$ and $\bar{u}^\gamma$, respectively.

To prove continuity of $\bar{u}^\bullet$ on $(0,\infty)$, we need several results. For readability, let

$$\beta(u) := \frac{1}{2}\sum_{i=1}^{n}\int_\Omega (-u_i)_+^2 dx + \frac{1}{2}\int_\Omega \left|\mathbf{1}^\top u - 1\right|^2 dx.$$

By definition of a global optimum, we have

$$\frac{1}{2}\|\bar{u}^\gamma - \varphi\|_{\mathbb{V}}^2 + \gamma\beta(\bar{u}^\gamma) \le \frac{1}{2}\|u-\varphi\|_{\mathbb{V}}^2 + \gamma\beta(u),\ \forall u \in \mathbb{V}.$$

Therefore, it follows from the feasibility of $\bar{u}$ and non-negativity of the objectives that

$$0 \le \theta(\gamma) \le \frac{1}{2}\|\bar{u}-\varphi\|_{\mathbb{V}}^2 = \theta_\infty,\ \forall \gamma > 0. \tag{1}$$

Furthermore, since $\beta(\cdot) \ge 0$, we deduce the uniform boundedness in $H^1(\Omega)^n$ of the set $\{\bar{u}^\gamma\}_{\gamma>0}$.

Let $\gamma_1, \gamma_2 \in (0,\infty)$ and define the associated optimal solutions of (P$_\gamma$) by $\bar{u}^{\gamma_i}$, $i = 1, 2$. Since $\beta : \mathbb{V} \to \mathbb{R}$ is continuous and convex, it is subdifferentiable. It follows then from the first-order optimality conditions for (P$_\gamma$) that

$$0 \in \partial(\frac{1}{2}\|\cdot-\varphi\|_{\mathbb{V}}^2+\gamma_i\beta(\cdot))(\bar{u}^{\gamma_i}) = \partial(\frac{1}{2}\|\cdot-\varphi\|_{\mathbb{V}}^2)(\bar{u}^{\gamma_i})+\gamma_i\partial\beta(\cdot)(\bar{u}^{\gamma_i}) \Longrightarrow -v_i \in \gamma_i\partial\beta(\cdot)(\bar{u}^{\gamma_i}),$$

where $v_i \in \partial(\frac{1}{2}\|\cdot-\varphi\|_{\mathbb{V}}^2)(\bar{u}^{\gamma_i}) = A(\bar{u}^{\gamma_i}-\varphi)$ and A is the (uniformly elliptic) bounded linear operator associated with the inner product on $\mathbb{V}$. Note that due to convexity and continuity, the sum rule for convex subdifferentials holds, cf. [17]. By definition of the subdifferential, we now have

$$\gamma_i\beta(z) \ge \gamma_i\beta(\bar{u}^{\gamma_i}) + \langle -v_i, z - \bar{u}^{\gamma_i}\rangle,\ \forall z \in \mathbb{V},\ i = 1,2. \tag{2}$$

For $i = 1$, we set $z = \bar{u}^{\gamma_2}$ and for $i = 2$, we set $z = \bar{u}^{\gamma_1}$ in (2). Adding the resulting terms yields

$$\gamma_1\beta(\bar{u}^{\gamma_2}) + \gamma_2\beta(\bar{u}^{\gamma_1}) \ge \gamma_1\beta(\bar{u}^{\gamma_1}) + \gamma_2\beta(\bar{u}^{\gamma_2}) + \langle -v_1, \bar{u}^{\gamma_2} - \bar{u}^{\gamma_1}\rangle + \langle -v_2, \bar{u}^{\gamma_1} - \bar{u}^{\gamma_2}\rangle.$$

Hence, we have

$$|\gamma_1-\gamma_2||\beta(\bar{u}^{\gamma_1})-\beta(\bar{u}^{\gamma_2})| \geq (\gamma_1-\gamma_2)(\beta(\bar{u}^{\gamma_2})-\beta(\bar{u}^{\gamma_1})) \geq \langle A(\bar{u}^{\gamma_1}-\bar{u}^{\gamma_2}), \bar{u}^{\gamma_1}-\bar{u}^{\gamma_2}\rangle = \|\bar{u}^{\gamma_1}-\bar{u}^{\gamma_2}\|_{\mathbb{V}}^2$$

Since $\{\beta(\bar{u}^{\gamma})\}_{\gamma>0}$ is bounded, there exists some $M > 0$ such that

$$\|\bar{u}^{\gamma_1} - \bar{u}^{\gamma_2}\|_{\mathbb{V}} \leq M^{1/2}|\gamma_1 - \gamma_2|^{1/2}, \forall \gamma_1, \gamma_2 > 0,$$

i.e., $\bar{u}^{\bullet} : (0,\infty) \to \mathbb{V}$ is Hölder continuous. Since β is convex and continuous, it is locally Lipschitz on any open convex set containing $\{\bar{u}^{\gamma}\}_{\gamma>0}$. Therefore, if we fix some $\gamma_0 \in (0,\infty)$ and consider $\bar{u}^{\bullet} : (\gamma_0-\varepsilon, \gamma_0+\varepsilon) \to \mathbb{V}$ for sufficiently small $\varepsilon > 0$, then $\bar{u}^{\bullet} : (\gamma_0-\varepsilon, \gamma_0+\varepsilon) \to \mathbb{V}$ is Lipschitz continuous. This proves 2.

Next, let $\gamma_k \to +\infty$ and define the sequence $\left\{\bar{u}^k\right\}$ by $\bar{u}^k := \bar{u}^{\gamma_k}$. Since $\left\{\bar{u}^k\right\} \subset \{\bar{u}^{\gamma}\}_{\gamma>0}$, which is bounded, there exists a weakly convergent subsequence $\left\{\bar{u}^{k_l}\right\}$ with limit $\bar{u}^{\star}$. Returning to (1) in this setting and dividing both sides by γ_{k_l}, it follows from the weak-lower semicontinuity of β that $\beta(\bar{u}^{\star}) = 0$; from which it can be easily deduce that $\bar{u}^{\star} \in \mathcal{G}$. But then given

$$\frac{1}{2}\|\bar{u}^{k_l} - \varphi\|_{\mathbb{V}}^2 \leq \frac{1}{2}\|\bar{u}^{k_l} - \varphi\|_{\mathbb{V}}^2 + \gamma_{k_l}\beta(\bar{u}^{k_l}) \leq \frac{1}{2}\|\bar{u} - \varphi\|_{\mathbb{V}}^2, \tag{3}$$

it follows from the weak lower-semicontinuity of $\|\cdot - \varphi\|_{\mathbb{V}}^2$ and the uniqueness of the global minimizer $\bar{u}$ that $\bar{u}^{\star} = \bar{u}$. Moreover, the uniqueness of $\bar{u}$ also implies, by the Urysohn property, that the entire sequence $\left\{\bar{u}^k\right\}$ converges weakly to $\bar{u}$. Using the previous inequality, we have once again by weak lower-semicontinuity and optimality (respectively):

$$\liminf_{k\to+\infty} \frac{1}{2}\|\bar{u}^k - \varphi\|_{\mathbb{V}}^2 \geq \frac{1}{2}\|\bar{u} - \varphi\|_{\mathbb{V}}^2 \qquad \text{and} \qquad \limsup_{k\to+\infty} \frac{1}{2}\|\bar{u}^k - \varphi\|_{\mathbb{V}}^2 \leq \frac{1}{2}\|\bar{u} - \varphi\|_{\mathbb{V}}^2.$$

Thus, $\|\bar{u}^k - \varphi\|_{\mathbb{V}}^2 \to \|\bar{u} - \varphi\|_{\mathbb{V}}^2$ and subsequently $\|\bar{u}^k\|_{\mathbb{V}} \to \|\bar{u}\|_{\mathbb{V}}$. Since $\mathbb{V}$ is a Hilbert space, it follows that $\bar{u}^k \to \bar{u}$ strongly in $\mathbb{V}$. This proves 3.

Finally, we prove 4. Since $0 \leq \theta(\gamma) \leq \gamma\beta(\varphi) \to 0$ as $\gamma \downarrow 0$, we obtain $\lim_{\gamma\downarrow 0}\theta(\gamma) = 0$. Moreover, local Lipschitz continuity of $\theta : (0,\infty) \to \mathbb{R}$ follows from that of $\bar{u}^{\bullet} : (0,\infty) \to \mathbb{V}$. To see that θ is monotonically increasing, fix $\gamma, \eta > 0$ consider that

$$\begin{aligned}\theta(\gamma+\eta) - \theta(\gamma) &\geq \frac{1}{2}\|\bar{u}^{\gamma+\eta} - \varphi\|_{\mathbb{V}}^2 + (\gamma+\eta)\beta(\bar{u}^{\gamma+\eta}) - (\frac{1}{2}\|\bar{u}^{\gamma+\eta} - \varphi\|_{\mathbb{V}}^2 + \gamma\beta(\bar{u}^{\gamma+\eta})) \\ &= \eta\beta(\bar{u}^{\gamma+\eta}) \geq 0.\end{aligned} \tag{4}$$

This completes the proof. □

Building on the results of Theorem 2.1, we show that $\theta : (0,\infty) \to \mathbb{R}$ is continuously differentiable.

Theorem 2.2. *Under the standing assumptions the optimal value function $\theta : (0,\infty) \to \mathbb{R}$ is continuously differentiable. The gradient of θ at $\bar{\gamma} > 0$ is given by*

$$\frac{\partial\theta}{\partial\gamma}(\bar{\gamma}) = \beta(\bar{u}^{\bar{\gamma}}).$$

Proof. Fix $\gamma, \eta > 0$. According to (4), we have

$$\eta^{-1}(\theta(\gamma+\eta) - \theta(\gamma)) \geq \beta(\bar{u}^{\gamma+\eta}). \tag{5}$$

Similarly, we have (for sufficiently small $\eta > 0$)

$$\theta(\gamma-\eta)-\theta(\gamma) \geq \frac{1}{2}\|\bar{u}^{\gamma-\eta}-\varphi\|_{\mathbb{V}}^2+(\gamma-\eta)\beta(\bar{u}^{\gamma-\eta})-(\frac{1}{2}\|\bar{u}^{\gamma-\eta}-\varphi\|_{\mathbb{V}}^2+\gamma\beta(\bar{u}^{\gamma-\eta})) = -\eta\beta(\bar{u}^{\gamma-\eta}).$$

Hence,

$$\eta^{-1}(\theta(\gamma-\eta) - \theta(\gamma)) \geq -\beta(\bar{u}^{\gamma-\eta}). \tag{6}$$

We also have the upper bounds

$$\theta(\gamma+\eta) - \theta(\gamma) \leq \frac{1}{2}\|\bar{u}^{\gamma}-\varphi\|_{\mathbb{V}}^2 + (\gamma+\eta)\beta(\bar{u}^{\gamma}) - (\frac{1}{2}\|\bar{u}^{\gamma}-\varphi\|_{\mathbb{V}}^2 + \gamma\beta(\bar{u}^{\gamma})) = \eta\beta(\bar{u}^{\gamma})$$

and (for sufficiently small $\eta > 0$)

$$\theta(\gamma-\eta) - \theta(\gamma) \leq \frac{1}{2}\|\bar{u}^{\gamma}-\varphi\|_{\mathbb{V}}^2 + (\gamma-\eta)\beta(\bar{u}^{\gamma}) - (\frac{1}{2}\|\bar{u}^{\gamma}-\varphi\|_{\mathbb{V}}^2 + \gamma\beta(\bar{u}^{\gamma})) = -\eta\beta(\bar{u}^{\gamma}).$$

This yields

$$\eta^{-1}(\theta(\gamma+\eta) - \theta(\gamma)) \leq \beta(\bar{u}^{\gamma}) \tag{7}$$
$$\eta^{-1}(\theta(\gamma-\eta) - \theta(\gamma)) \leq -\beta(\bar{u}^{\gamma}). \tag{8}$$

It follows from Theorem 2.1.2, the continuity of β, (5), and (7) that

$$\lim_{\eta\downarrow 0} \eta^{-1}(\theta(\gamma+\eta) - \theta(\gamma)) = \beta(\bar{u}^{\gamma}).$$

Similarly, we have from Theorem 2.1.2., the continuity of β, (6), and (8) that

$$\lim_{\eta\downarrow 0} \eta^{-1}(\theta(\gamma-\eta) - \theta(\gamma)) = -\beta(\bar{u}^{\gamma}).$$

Hence, θ is differentiable from $(0,\infty) \to \mathbb{R}$. Finally, as $\bar{u}^{\bullet} : (0,\infty) \to \mathbb{V}$ and β are continuous, θ is continuously differentiable. □

Remark 2.3 (Relation to Danskin's Theorem)**.** *Since β was kept rather general in Theorems 2.1 and 2.2, it appears that the value function is continuously differentiable for a wide variety of constraints beyond our setting. Note that Theorem 2.2 is not a direct consequence of Danskin's theorem. Indeed, the hypotheses of Danskin's theorem (cf. [6, Prop. 4.12]) require, in particular, that the objective functional is continuous and $\inf$-compact on some Hausdorff topological vector space. In our case, the level sets of the objective are $\inf$-compact in the weak topology on $\mathbb{V}$. However, the objective is weakly lower-semicontinuous, but not continuous with respect to the weak topology. Therefore, we cannot simply embed the problem into the setting of Danskin's theorem by replacing $(\mathbb{V}, \|\cdot\|_{\mathbb{V}})$ by $(\mathbb{V}, \sigma_{\text{weak}})$.*

In order to approximate θ in the path-following scheme by a simple model function, we need second-order information. This will require a differential sensitivity analysis of $\bar{u}^\bullet$. For this discussion, we will make use of the first-order necessary and sufficient optimality conditions for (P_γ), which for each $i = 1, \dots, n$ are given by

$$-\Delta \bar{u}_i^\gamma + \bar{u}_i^\gamma - \lambda_i^\gamma + \mathbf{1}\mu^\gamma = -\Delta\varphi_i + \varphi_i, \quad \text{in } \Omega, \tag{9a}$$

$$\frac{\partial(\bar{u}_i^\gamma - \varphi_i)}{\partial n} = 0, \quad \text{on } \Gamma, \tag{9b}$$

$$\lambda_i^\gamma = \gamma(-\bar{u}_i^\gamma)_+, \tag{9c}$$

$$\mu^\gamma = \gamma(\mathbf{1}^\top \bar{u}^\gamma - 1). \tag{9d}$$

Note that $-\lambda_i^\gamma = \gamma\partial(\frac{1}{2}\|(-\cdot)_+\|^2_{L^2(\Omega)})(\bar{u}_i^\gamma)$ and $\mu^\gamma = \gamma\partial_{u_i}(\frac{1}{2}\|\mathbf{1}^\top \cdot -1\|^2_{L^2(\Omega)})(\bar{u}^\gamma)$. In addition, we define the difference quotients:

$$d_i^\eta := \frac{\bar{u}_i^{\gamma+\eta} - \bar{u}_i^\gamma}{\eta}, \quad \chi_i^\eta := \frac{\lambda_i^{\gamma+\eta} - \lambda_i^\gamma}{\eta}, \quad \nu^\eta := \frac{\mu^{\gamma+\eta} - \mu^\gamma}{\eta}.$$

We then have the following result.

Proposition 2.4. *The directional derivative of the composition $(\beta \circ \bar{u}^\bullet) : (0, \infty) \to \mathbb{R}$ at $\gamma > 0$ in direction 1 is given by*

$$\beta'(\bar{u}^\gamma; 1) = -\sum_{i=1}^n \int_\Omega (-\bar{u}_i^\gamma)_+ \bar{d}_i \, dx + \int_\Omega (\mathbf{1}^T \bar{u}^\gamma - 1)\bar{d}_i \, dx,$$

where $\bar{d} \in \mathbb{V}$ is the unique solution of the following variational problem:

$$\min \Big\{ \frac{1}{2}\|d\|^2_{\mathbb{V}} - \sum_{i=1}^n \big((-\bar{u}_i^\gamma)_+ - (\mathbf{1}^\top \bar{u}^\gamma - 1), d_i\big)_{L^2(\Omega)} + \frac{\gamma}{2}\int_\Omega \left|\mathbf{1}^\top d\right|^2 dx + \\ \frac{\gamma}{2}\sum_{i=1}^n \int_{\{\bar{u}_i^\gamma = 0\}} (-d_i)_+^2 dx + \frac{\gamma}{2}\int_{\{\bar{u}_i^\gamma < 0\}} |d_i|^2 dx \text{ over } d \in \mathbb{V} \Big\}. \tag{D_γ}$$

Moreover, it holds that

$$\beta'(\bar{u}^\gamma; 1) \leq 0, \ \forall \gamma > 0.$$

Hence, $\theta : (0, \infty) \to \mathbb{R}$ is concave.

Proof. We first show that $\{d^\eta\}_{\eta > 0}$ is uniformly bounded in $\mathbb{V}$. First note that d_i^η solves the following elliptic PDE with homogeneous Neumann boundary conditions:

$$-\Delta d_i^\eta + d_i^\eta - \chi_i^\eta + \nu^\eta = 0, \quad \text{in } \Omega, \tag{10a}$$

$$\frac{\partial d_i^\eta}{\partial n} = 0, \quad \text{on } \Gamma. \tag{10b}$$

Now note that $\sum_{i=1}^n \|d_i^\eta\|_V^2 = \sum_{i=1}^n \langle -\Delta d_i^\eta + d_i^\eta, d_i^\eta\rangle$ and by the monotonicity of the gradients of β that

$$\sum_{i=1}^n \|d_i^\eta\|_V^2 \leq \sum_{i=1}^n \langle -\Delta d_i^\eta + d_i^\eta, d_i^\eta\rangle - \sum_{i=1}^n \gamma\eta^{-2}((-\bar{u}_i^{\gamma+\eta})_+ - (-\bar{u}_i^\gamma)_+, \bar{u}_i^{\gamma+\eta} - \bar{u}_i^\gamma)_{L^2(\Omega)} + \sum_{i=1}^n \gamma\eta^{-2}((\mathbf{1}^\top \bar{u}^{\gamma+\eta} - 1 - (\mathbf{1}^\top \bar{u}^\gamma - 1), \bar{u}_i^{\gamma+\eta} - \bar{u}_i^\gamma)_{L^2(\Omega)}$$

Then, testing (10a) with d_i^η and summing over i we have

$$\sum_{i=1}^n \langle -\Delta d_i^\eta + d_i^\eta, d_i^\eta\rangle = \sum_{i=1}^n \langle \chi_i^\eta, d_i^\eta\rangle - \sum_{i=1}^n \langle \nu^\eta, d_i^\eta\rangle.$$

Next, given

$$\begin{aligned}\sum_{i=1}^n \langle \chi_i^\eta, d_i^\eta\rangle - \sum_{i=1}^n \langle \nu^\eta, d_i^\eta\rangle - \sum_{i=1}^n \gamma\eta^{-2}((-\bar{u}_i^{\gamma+\eta})_+ - (-\bar{u}_i^\gamma)_+, \bar{u}_i^{\gamma+\eta} - \bar{u}_i^\gamma)_{L^2(\Omega)} + \\ \sum_{i=1}^n \gamma\eta^{-2}((\mathbf{1}^\top \bar{u}^{\gamma+\eta} - 1 - (\mathbf{1}^\top \bar{u}^\gamma - 1), \bar{u}_i^{\gamma+\eta} - \bar{u}_i^\gamma)_{L^2(\Omega)} \\ = \sum_{i=1}^n ((-\bar{u}_i^{\gamma+\eta})_+, d_i^\eta)_{L^2(\Omega)} - \sum_{i=1}^n ((\mathbf{1}^\top \bar{u}^{\gamma+\eta} - 1, d_i^\eta)_{L^2(\Omega)},\end{aligned}$$

we have

$$\sum_{i=1}^n \|d_i^\eta\|_V^2 \leq \sum_{i=1}^n ((-\bar{u}_i^{\gamma+\eta})_+, d_i^\eta)_{L^2(\Omega)} - \sum_{i=1}^n ((\mathbf{1}^\top \bar{u}^{\gamma+\eta} - 1, d_i^\eta)_{L^2(\Omega)}. \tag{11}$$

It follows from Theorem 2.1.3, that $\{d^\eta\}_{\eta>0}$ is uniformly bounded in $\mathbb{V}$. Next, we derive expansions for λ_i^η and μ^η.

From the Lebesgue dominated convergence theorem we see that the superposition operator $(\cdot)_+ : L^2(\Omega) \to L^2(\Omega)$ is Lipschitz continuous and Hadamard directionally differentiable with

$$(\cdot)'_+(\bar{u}_i^\gamma; d_i^\eta)(x) = \begin{cases} d_i^\eta(x) & \text{on } \{\bar{u}_i^\eta > 0\} \cup \{\bar{u}_i^\eta = 0, d_i^\eta > 0\}, \\ 0 & \text{otherwise.}\end{cases} \tag{12}$$

Moreover, due to [21, Proposition 3.1] it is continuous in the second variable. Since $\{d_i^\eta\}_{\eta>0}$ is uniformly bounded in V and

$$\chi_i^\eta = \eta^{-1}((\gamma+\eta)(-\bar{u}_i^{\gamma+\eta})_+ - \gamma(-\bar{u}_i^\gamma)_+) = \gamma\left(\frac{(-(\bar{u}_i^\gamma + \eta d_i^\eta))_+ - \gamma(-\bar{u}_i^\gamma)_+)}{\eta}\right) + (-\bar{u}_i^{\gamma+\eta})_+,$$

there exists a small o-function such that

$$\chi_i^\eta = \gamma(\cdot)'_+(-\bar{u}_i^\gamma; -d_i^\eta) + o(1) + (-\bar{u}_i^{\gamma+\eta})_+. \tag{13}$$

Arguing similarly for ν^η, we have

$$\nu^\eta = \gamma \mathbf{1}^\top d^\eta + \mathbf{1}^\top \bar{u}^{\gamma+\eta} - 1. \tag{14}$$

Now, letting $\eta_k \downarrow 0$, there exists a weakly converging subsequence $\left\{d_i^{k_l}\right\}$ with $d_i^{k_l} := d_i^{\eta_{k_l}}$ and a weak limit point $\bar{d}_i \in V$. Since V embeds compactly into $L^2(\Omega)$ and $\bar{u}^\bullet : (0,\infty) \to \mathbb{V}$ is continuous, we have

$$\begin{aligned}
&\chi_i^{\eta_{k_l}} \to \gamma(\cdot)'_+(-\bar{u}_i^\gamma; -\bar{d}_i) + (-\bar{u}_i^\gamma)_+ \text{ strongly in } L^2(\Omega)\\
&\nu^{\eta_{k_l}} \to \gamma \mathbf{1}^\top \bar{d} + \mathbf{1}^\top \bar{u}^\gamma - 1 \text{ strongly in } L^2(\Omega)
\end{aligned}$$

Moreover, $\bar{d}_i$ solves

$$-\Delta \bar{d}_i + \bar{d}_i - \gamma(\cdot)'_+(-\bar{u}_i^\gamma; -\bar{d}_i) + \gamma \mathbf{1}^\top \bar{d} = (-\bar{u}_i^\gamma)_+ - (\mathbf{1}^\top \bar{u}^\gamma - 1), \quad \text{in } \Omega, \tag{15a}$$

$$\frac{\partial \bar{d}_i}{\partial n} = 0, \quad \text{on } \Gamma. \tag{15b}$$

It is then easy to see that the vector $\bar{d} \in \mathbb{V}$ solves (D_γ). Since (D_γ) has a unique solution, we can deduce by the Urysohn property that the entire sequence $\left\{d^k\right\}$ converges weakly to $\bar{d}$.

We are now ready to derive a formula for the directional derivative of $\theta'(\gamma)$. Consider that

$$\begin{aligned}
\eta^{-1}(\beta(\bar{u}^{\gamma+\eta}) - \beta(\bar{u}^\gamma)) = (2\eta)^{-1} \sum_{i=1}^n \int_\Omega [(-\bar{u}_i^{\gamma+\eta})_+^2 - \\
(-\bar{u}_i^\gamma)_+^2] dx + (2\eta)^{-1} \int_\Omega [|\mathbf{1}^T \bar{u}^{\gamma+\eta} - 1|^2 - |\mathbf{1}^T \bar{u}^\gamma - 1|^2] dx.
\end{aligned}$$

Using the chain rule, formula (12) and $\bar{u}^{\gamma+\eta} = \bar{u}^\gamma + \eta d^\gamma$, we have

$$\eta^{-1}(\beta(\bar{u}^{\gamma+\eta}) - \beta(\bar{u}^\gamma)) = \sum_{i=1}^n \int_\Omega [-(-\bar{u}_i^\gamma)_+ d_i^\eta] dx + \int_\Omega [(\mathbf{1}^T \bar{u}^\gamma - 1) d_i^\eta] dx + o(1).$$

Since $\eta_k \to 0$ implies $d^{\eta_k} \rightharpoonup \bar{d}$ (weakly in $\mathbb{V}$), we can use the compactness of the embedding V into $L^2(\Omega)$ and pass to the limit in the previous equation. This yields

$$\beta'(\bar{u}^\gamma; 1) = -\sum_{i=1}^n \int_\Omega (-\bar{u}_i^\gamma)_+ \bar{d}_i \, dx + \int_\Omega (\mathbf{1}^T \bar{u}^\gamma - 1) \bar{d}_i \, dx.$$

Finally, it follows from (11) that $\beta'(u^\gamma; 1) \leq 0$. This completes the proof. □

3 Solving the First-Order System

3.1 A Semismooth Newton Iteration

The main component of the algorithm involves the direct solution of (9) using a semismooth Newton method. We refer the reader to the papers [23, 14] for the full theory of this method and

only note that in the current setting, the iterates are generated by solving the following (reduced) linear system: Given $\gamma > 0$ and u_{old}, solve

$$A(u-\varphi) + \gamma I_{\{u_{\mathrm{old}}<0\}} u + \gamma \mathbf{1}(\mathbf{1}^\top u - 1) = 0 \text{ in } \Omega, \tag{16a}$$

$$\frac{\partial(u-\varphi)}{\partial n} = 0 \text{ on } \partial\Omega \tag{16b}$$

for u_{new}. Here, A is the PDE-operator associated with the $\mathbb{V}$-inner product and $I_{\{u_{\mathrm{old}}<0\}}$ is the characteristic function for the set $\{u_{\mathrm{old}} < 0\}$. Once the residual of (9) is sufficiently small, the method terminates and γ is updated. The previous γ-dependent solution serves as the starting value for the next iteration. We summarize this procedure in Algorithm 3.1.

Algorithm 3.1 H^1-Projection onto the Gibbs Simplex

Input: tol > 0, $\gamma > 0$, $\varphi \in H^1(\Omega, \mathbb{R}^n)$, $k \leftarrow 0$, $u_0 \leftarrow \frac{1}{n}\mathbf{1}$

repeat ▷ Outer loop
  $k \leftarrow k+1$, $l \leftarrow 1$, $u_{k,1} \leftarrow u_{k-1}$
  repeat ▷ Inner loop
    Set $\gamma \leftarrow \gamma_k$, $u_{\mathrm{old}} \leftarrow u_{k,l}$ and solve (16) for $u_{k,l+1}$
    $l \leftarrow l+1$
  until Stopping criterion is satisfied
  $u_k \leftarrow u_{k,l+1}$
  $\lambda_k \leftarrow -\gamma_k(-u_k)_+$, $\mu_k \leftarrow \gamma_k(\mathbf{1}^\top u_k - 1)$
  Update γ_{k+1}
until Stopping criterion is satisfied

Defining the residuals

$$r_k^1 = \|(-u_k)_+\|_{L^2(\Omega,\mathbb{R}^n)},\ r_k^2 = \|\mathbf{1}^\top u_k - 1\|_{L^2(\Omega)},\ r_k^3 = (\lambda_k, u_k)_{L^2(\Omega)},$$

the outer loop terminates whenever

$$\|(r_k^1, r_k^2, r_k^3)\|_2 \leq \mathrm{tol}.$$

Note that r_k^1, r_k^2 are feasibility errors, whereas r_k^1, r_k^3 represent the complementarity error. The stopping criteria for the inner loop depend on the type of γ-update strategy. These are detailed below in Section 4.

3.2 Feasibility Restoration

In some applications, e.g., topology optimization, where the phase field functions $u_1, \ldots, u_n$ arise as parameters in a second-order PDE-operator, it is imperative that $u \in \mathcal{G}$. For such situations, we are forced to employ a mesh-dependent numerical method. However, the (very) warm start provided by Algorithm 3.1 provides a point that is sufficiently close to the true solution of (P) so that the jump to feasibility requires minimal effort. We demonstrate this fact in the examples.

4 Path-Following Scheme

We now detail a strategy for efficiently updating γ_k in Algorithm 3.1. As suggested in [16], an ideal update scheme for γ (at step k) would be to choose γ_{k+1} such that

$$|\theta_\infty - \theta(\gamma_{k+1})| \leq \tau_k |\theta_\infty - \theta(\gamma_k)|, \tag{17}$$

where $\tau_k \downarrow 0$ is a given null sequence. However, as noted there, this is not a practical strategy as we cannot evaluate $\theta(\gamma_{k+1})$ and θ_∞. Nevertheless, our analysis shows that $\theta : (0, \infty) \to \mathbb{R}$ is positive, bounded, increasing, continuously differentiable and concave. Hence, we first fit a model function, using the information available at step k, and then update according to an approximation of (17). This serves the basis for the path-following approach.

Our approach follows the arguments in [16, Sections 5-6]. We therefore consider the following model function $m : (0, \infty) \to \mathbb{R}$ given by

$$m(\delta) := C_1 - \frac{C_2}{E + \delta}$$

as an approximation of θ. Given $\gamma > 0$ and $\bar{u}^\gamma$, we determine the (γ-dependent) constants in each of the model functions by setting

$$m(0) = 0, \quad m(\gamma) = \theta(\gamma), \quad m'(\gamma) = \theta'(\gamma). \tag{19}$$

This amounts to:

$$\begin{aligned} E &= \gamma^2 \beta(\bar{u}^\gamma)(\frac{1}{2}\|\bar{u}^\gamma - \varphi\|_{\mathbb{V}}^2)^{-1}, \\ C_2 &= \gamma^{-1} E(E + \gamma)(\frac{1}{2}\|\bar{u}^\gamma - \varphi\|_{\mathbb{V}}^2 + \gamma\beta(\bar{u}^\gamma)), \\ C_1 &= C_2 E^{-1}. \end{aligned}$$

Clearly, $E, C_1, C_2 > 0$. Moreover, it can be argued using (3) that $\gamma\beta(\bar{u}^\gamma) \to 0$ as $\gamma \to +\infty$. Hence, $m(\cdot)$ monotonically increasing, $m(\cdot) \leq C_1$ and

$$C_1 = \left(\frac{\gamma\beta(\bar{u}^\gamma)}{\frac{1}{2}\|\bar{u}^\gamma - \varphi\|_{\mathbb{V}}^2} + 1 \right) \theta(\gamma)$$

implies that $m(\delta) \to C_1$ as $\delta \to \infty$ and $C_1 \to \theta_\infty$ as $\gamma \to \infty$. Now, we may use (17), which leads to the following explicit update formula:

$$\gamma_{k+1} = \frac{C_{2,k}}{\tau_k |C_{1,k} - \theta(\gamma_k)|} - E_k. \tag{21}$$

By choosing three different reference values for γ in (19), we see in Figure 1 that m is an excellent local approximation of θ.

In this method, the Newton iteration (inner loop) is terminated whenever the active index sets $\{u_k < 0\}$ are the same in two subsequent iterations or

$$\|A(u_{k,l+1} - \varphi) - \gamma_k(-u_{k,l+1})_+ + \gamma_k \mathbf{1}(\mathbf{1}^\top u_{k,l+1} - 1)\|_{H^1(\Omega,\mathbb{R}^n)^*} \leq \mathrm{tol}. \tag{22}$$

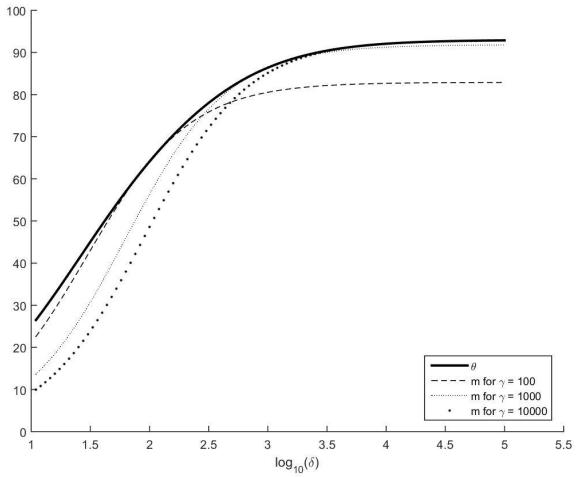

Figure 1: Approximations m of θ taken at different reference points γ

5 Numerical results

In this section, we demonstrate the performance and wide applicability of the proposed methods on three examples. The first example shows the mesh-independent behavior of the path-following methods and compares this to the first-discrete-then-optimize approach suggested above, which utilizes the primal-dual active set method (adapted to our setting) from [14]. This method does not admit a function space convergence analysis and, hence, is expected to behave mesh dependently, i.e., utilizing the same stopping rule and initial point on all meshes, the number of iterations until successful termination will grow as the mesh is refined. For the latter two examples, we consider two optimization problems from mathematical image processing and labeled data classification that may be solved using a gradient-flow or projected gradient method, see e.g. [3]. At each step of the line search in the projected gradient method, a projection onto the Gibbs simplex is required.

In all examples, the underlying domain $\Omega = (0,1)^2$ is discretized using a uniform triangular mesh with parameter $h > 0$. The function space $H^1(\Omega)$ is discretized using the associated nodal basis comprised of the standard continuous piecewise affine finite elements, see, e.g., [10, 7]. Note that the pointwise almost-everywhere constraints are approximated in the finite dimensional problem by requiring the coefficients for each nodal basis function to be in the finite dimensional Gibbs simplex. In order to calculate the $H^1(\Omega;\mathbb{R}^n)^*$-norm in (22), we need the Riesz representation of the dual object. This is done by solving a (vector-valued) Poisson problem with homogeneous Neumann boundary conditions with the residual of interest on the right-hand side. We make use of the P1AFEM package [11]. Concerning the path-following parameters, we choose $\tau_k = 10^{-2k}$, initial parameter $\gamma_1 = 10^4$, maximum parameter $\gamma_{\max} = 10^{14}$ and the stopping criterion tol $= 10^{-6}$. PDAS is stopped once the residual of the optimality conditions to (P) drops to 10^{-10}.

5.1 Projection onto the Gibbs Simplex

For this example, we set $n = 3$ and construct φ_i such that the optimal solution of (P) exhibits biactivity at the optimal solution, i.e, both the inequality constraint and the associated Lagrange multiplier are equal to zero. Such problems can be notoriously difficult to solve. Let

$$\begin{aligned}
\hat{v}_1(x,y) &= \begin{cases} 0 & \text{if } x \in (0,1),\ y \in (0,\frac{1}{2}) \\ \sin(y-\frac{1}{2})\cos(xy) & \text{if } x \in (0,1),\ y \in [\frac{1}{2},1) \end{cases} \\
\hat{v}_2(x,y) &= 2 + \cos(10xy),\ (x,y) \in (0,1)^2 \\
\hat{v}_3(x,y) &= 0,\ (x,y) \in (0,1)^2
\end{aligned}$$

and then set

$$v_i = \frac{\hat{v}_i}{\hat{v}_1 + \hat{v}_2 + \hat{v}_3}.$$

for $i = 1, 2, 3$. Note that $v = (v_1, v_2, v_3)$ satisfies the constraints in $\mathcal{G}$. Similarly, we define the functions

$$\begin{aligned}
\lambda_1(x,y) &= \begin{cases} 1 & \text{if } x \in (0,1),\ y \in (0,\frac{1}{2}) \\ 0 & \text{if } x \in (0,1),\ y \in [\frac{1}{2},1) \end{cases} \\
\lambda_2(x,y) &= 0,\ (x,y) \in (0,1)^2 \\
\lambda_3(x,y) &= 0,\ (x,y) \in (0,1)^2, \\
\mu(x,y) &= 1,\ (x,y) \in (0,1)^2
\end{aligned}$$

and construct φ by solving the problem:

$$\begin{aligned}
-\Delta\varphi + \varphi &= -\Delta v + v + \lambda - \mathbf{1}\mu, \quad \text{in } \Omega, \\
\frac{\partial(\varphi - v)}{\partial n} &= 0, \quad \text{on } \Gamma.
\end{aligned}$$

Note that $\lambda = (\lambda_1, \lambda_2, \lambda_3) \in L^2(\Omega)^3$, that $(\lambda, v)_{L^2} = 0$ and that the above system forms the Karush-Kuhn-Tucker conditions for (P). Thus, v is its optimal solution, λ is the Lagrange multiplier associated with the inequality constraint and biactivity is present on the whole third component. The starting point was chosen as φ.

The results of Algorithm 3.1 are presented in Table 1. We compare the behavior of the primal-dual active set (PDAS) and the path-following method (PF) on six different meshes with decreasing h. For PDAS, 'total iter' refers to the number of linear solves. For PF, 'inner steps l' refers to the total number of linear solves used in the Newton iterations, 'outer steps k' is the number of γ updates, and 'feas. step' counts the number of linear solves to optimality needed when using the solution of PF as a starting point in the PDAS method. This qualitatively demonstrates how close the solution of the approximating problems gets to solving (P). Note that no adaptive mesh refinement or nested grid strategies, where one would consider subsequently refined meshes and use the prolongation of the solution on the coarser mesh as the starting point on the finer mesh, are considered. Clearly, the PF strategies behave much better over mesh refinements than PDAS. For example, with $h = 2^{-8}$, the true solution to (P) is obtained in 99 iterations with PDAS and 15 with PF.

h	2^{-4}	2^{-5}	2^{-6}	2^{-7}	2^{-8}	2^{-9}
PDAS: total iter	10	16	33	58	99	196
PF: inner steps l	12	12	14	14	15	14
outer steps k	3	3	3	3	3	3
feas. step	1	1	1	1	1	1

Table 1: Results for projection of φ onto the Gibbs simplex

5.2 Applications in Image Processing and Labeled Data Classification

In this subsection, we consider two examples inspired by the work in [2, 12]. Having n phases, the common goal is to minimize a criterion J such that the phases are (almost) pure. This amounts to solving the problem

$$\min_{u\in H^1(\Omega;\mathbb{R}^n)} \left\{ J(u) + \zeta \int_\Omega \left(\epsilon|\nabla u|^2 + \frac{1}{\epsilon} u(1-u) \right) dx \;\middle|\; u \in \mathcal{G} \right\}, \tag{23}$$

where $\zeta > 0$ is the scaling parameter and parameter $\epsilon > 0$ is proportional to the interfacial width. The second part of the objective is the multiphase Ginzburg-Landau energy, which along with the indicator function for $\mathcal{G}$, serves as a penalty functional that avoids pathological free boundaries between the phases. Together, the sum of these functionals Γ-converges to the perimeter of the characteristic functions associated with the true phases in a sharp interface regime [19, 20]. As such, straight edges are more favorable. We use the standard projected gradient method with line search to solve (23).

Labeled Data Classification
In the first example we separate $n = 3$ different classes of observations with known labels. First, we 'train' three segmentation phase field functions u on M noisy data and then test the accuracy of these solutions on further noisy realizations. This amounts to solving (23) with

$$J(u) = \sum_{m=1}^{M} \left(1 - \frac{1}{|B(x_m,\delta)|} \int_{B(x_m,\delta)} u_{j(m)} dx \right), \tag{24}$$

where $m \in \{1,\dots,M\}$ is the index of a noisy data point $x_m \in \Omega$ and $B(x_m,\delta)$ is a small ball of radius $\delta > 0$ around this point; $j(m) \in \{1,2,3\}$ refers to the known label (class) of this point. Due to the structure of the Gibbs simplex, J is minimized when $u_{j(m)} = 1$ and $u_i = 0$ for $i \neq j(m)$ on $B(x_m,\delta)$. Thus, J expressed the number of incorrectly classified points. In (23) we consider $\epsilon = 2^{-6}$ and $\zeta = 2\cdot 10^{-2}$ and in (24) we set $\delta = 3\cdot 2^{-7}$. The algorithm was started with a uniform mixture of phases.

We randomly generated $M = 1500$ points (500 for each category $+$/left half circle, $\circ$/bottom half circle, $\times$/right half circle), see the left-hand side of Figure 2. We then solved the optimization problem in order to train three phase field functions, see the right-hand side of Figure 2. In order to test the accuracy of these fields, we again generated the same number of points and checked how many of these points lie in the correct region. In this instance, the solution was correct for 99.6% of the data points. The numerical comparison of both methods is shown in Table 2. Even though PF needs more iteration than PDAS, the iteration growth is much slower upon mesh

refinement. The difference diminishes when we use warm start from the previous mesh. As we will see it the next application, on sufficiently fine meshes PF will eventually need less iterations than PDAS.

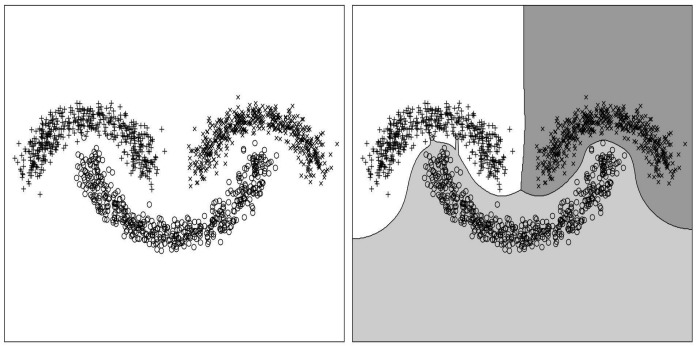

Figure 2: Classification results: Generated random points in the shape of three half-moons (left) and the obtained three regions used for categorization of points (right).

	Uniform phases start				Warm start			
h	2^{-6}	2^{-7}	2^{-8}	2^{-9}	2^{-6}	2^{-7}	2^{-8}	2^{-9}
PDAS: total iter	33	39	58	89	33	29	39	48
increase		18.18%	48.72%	53.45%		-12.12%	34.48%	23.08%
PF: inner steps l	59	68	85	113	59	51	68	81
increase		15.25%	25.00%	32.94%		-13.56%	33.33%	19.12%
feas. step	8	13	14	16	8	9	15	17

Table 2: Classification results: Number of iterations on uniform meshes with element size h and the increases with respect to previous iterations. The first part depicts the start with uniform phases, the other one the warm start from the previous mesh. The projected gradients needed 4 iterations with 7 projections in both cases.

Inpainting

For the inpainting application, we assume that an original image (top left of Figure 3) is degraded by adding a text on top of it to produce a known image $\hat{u}$ (top right of Figure 3).[1] The goal is then to recover the original image as faithfully as possible. We follow the suggestion in [2] to use a modified Allen-Cahn equation for image inpainting, which leads to (23) with

$$J(u) = \int_\Omega (u - \hat{u})^2 dx. \tag{25}$$

In (23) we consider $\epsilon = 2^{-7}$ and $\zeta = 7 \cdot 10^{-3}$. The algorithm was started with the degraded image.

On the contrary to the previous cases, we use non-uniform mesh, where the phase interface is much more refined, see the bottom right part of Figure 3 which depicts the mesh for the right panda eye. For simplicity, we consider a locally adapted mesh which is given. A similar mesh

[1] The picture can be downloaded from many wallpaper sites such as http://dairennsa.net/

may be obtained from developing an a posteriori error estimates for adaptive mesh refinement see e.g. [1] or [13]. As the latter technique, however, is not the focus of this work, we rely on an a priori mesh. For simplicity, we consider locally adapted mesh which is given a priori; for results The recovered image is shown in the bottom left part of Figure 3. The numerical evidence is summarized in Table 3, where h is the smallest element size. On the coarsest mesh PDAS outperforms PF but as the mesh is getting finer, PF needs less iterations to converge to the solution. In all cases, the projected gradients needed 12 iterations with 26 projections.

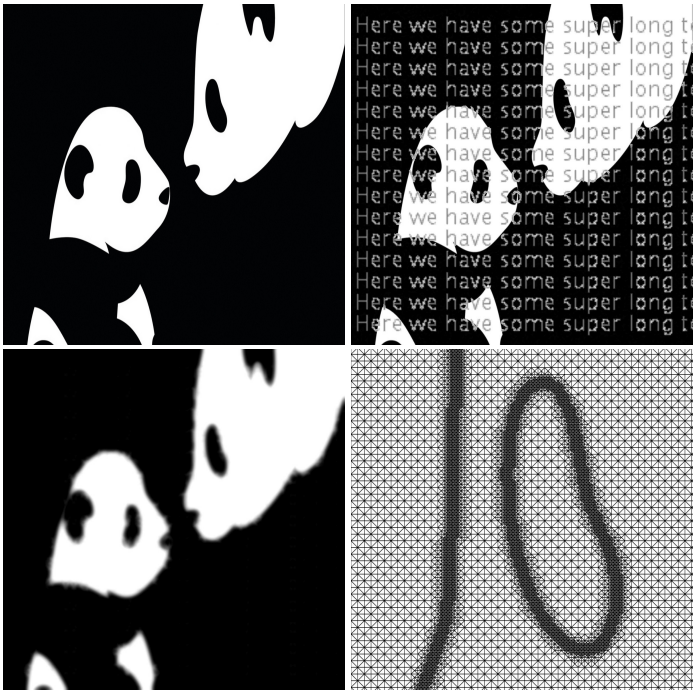

Figure 3: Inpainting results: Initial image (top left), devalued image (top right), recovered image (bottom left) and mesh detail of the right panda eye (bottom right).

h	2^{-8}	2^{-9}	2^{-10}	2^{-11}	2^{-11}	2^{-12}
Node number	66049	80500	110065	169733	289685	529971
PDAS: total iter	96	108	125	159	194	245
PF: inner steps	155	164	176	190	200	200
feas. step	35	33	45	48	51	74

Table 3: Inpainting results: Non-uniform mesh with refined mesh interface; h is the size of the smallest element. The projected gradients needed 12 iterations with 26 projections.

6 Summary

In this paper, we motivate the need for a function-space-based algorithm for the H^1-projection onto the multiphase Gibbs simplex. In particular, the wide array of applications alone involving optimization or control problems with the Gibbs simplex as a constraint warrants such a study.

In order to make the best use of fast second-order methods for semismooth equations, an analytical path-following study is carried out. Our analytical approach drastically reduces the length of the arguments and necessary assumptions required in [16, 15]; the two works that inspired our approach. Due to the generality of our arguments, they can clearly be extended to similar constraint sets. As in [16, 15], we suggest a path-following strategy for the update of the Moreau-Yosida regularization parameter, which is based on a concave model function. In our numerical experiments, we first demonstrate the clear advantage of our methods over a first-discretize-then-optimize approach. Finally, we consider two small examples inspired by recent work in [2, 12] that require the H^1-projection onto the Gibbs simplex. Further applications, for example in topology optimization will be considered in future work.

References

[1] L. Baňas and R. Nürnberg. A posteriori estimates for the Cahn-Hilliard equation with obstacle free energy. *ESAIM: Mathematical Modelling and Numerical Analysis*, 43(5):1003–1026, 2009.

[2] A. L. Bertozzi, S. Esedoglu, and A. Gillette. Inpainting of binary images using the Cahn-Hilliard equation. *IEEE Transactions on image processing*, 16(1):285–291, 2007.

[3] D. P. Bertsekas. On the Goldstein-Levitin-Polyak gradient projection method. *IEEE Trans. Automatic Control*, AC-21(2):174–184, 1976.

[4] L. Blank, M. H. Farshbaf-Shaker, H. Garcke, C. Rupprecht, and V. Styles. Multi-material phase field approach to structural topology optimization. In *Trends in PDE constrained optimization*, volume 165 of *Internat. Ser. Numer. Math.*, pages 231–246. Birkhäuser/Springer, Cham, 2014.

[5] L. Blank, H. Garcke, L. Sarbu, T. Srisupattarawanit, V. Styles, and A. Voigt. Phase-field approaches to structural topology optimization. In *Constrained Optimization and Optimal Control for Partial Differential Equations*, pages 245–256. Springer, 2012.

[6] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, 2000.

[7] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[8] H. Brézis and G. Stampacchia. Sur la régularité de la solution d'inéquations elliptiques. *Bulletin de la Société Mathématique de France*, 96:153–180, 1968.

[9] M. Burger, L. He, and C.-B. Schönlieb. Cahn-Hilliard inpainting and a generalization for grayvalue images. *SIAM Journal on Imaging Sciences*, 2(4):1129–1167, 2009.

[10] P. G. Ciarlet. *The finite element method for elliptic problems*. North-Holland Publishing Co., Amsterdam-New York-Oxford, 1978. Studies in Mathematics and its Applications, Vol. 4.

[11] S. Funken, D. Praetorius, and P. Wissgott. Efficient implementation of adaptive P1-FEM in Matlab. *Computational Methods in Applied Mathematics Comput. Methods Appl. Math.*, 11(4):460–490, 2011.

[12] C. Garcia-Cardona, E. Merkurjev, A. L. Bertozzi, A. Flenner, and A. G. Percus. Multiclass data segmentation using diffuse interface methods on graphs. *IEEE Transactions on pattern analysis and machine intelligence*, 36(8):1600–1613, 2014.

[13] M. Hintermüller, M. Hinze, and M. H. Tber. An adaptive finite-element Moreau-Yosida-based solver for a non-smooth Cahn-Hilliard problem. *Optimization Methods and Software*, 26(4-5):777–811, 2011.

[14] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semismooth Newton method. *SIAM Journal on Optimization*, 13(3):865–888, 2002.

[15] M. Hintermüller and K. Kunisch. Feasible and noninterior path-Řfollowing in constrained minimization with low multiplier regularity. *SIAM Journal on Control and Optimization*, 45(4):1198–1221, 2006.

[16] M. Hintermüller and K. Kunisch. Path-following methods for a class of constrained minimization problems in function space. *SIAM Journal on Optimization*, 17(1):159–187, 2006.

[17] A. D. Ioffe and V. M. Tihomirov. *Theory of extremal problems*, volume 6 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam-New York, 1979. Translated from the Russian by Karol Makowski.

[18] D. Kinderlehrer and G. Stampacchia. *An introduction to variational inequalities and their applications*, volume 31 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. Reprint of the 1980 original.

[19] L. Modica. The gradient theory of phase transitions and the minimal interface criterion. *Archive for Rational Mechanics and Analysis*, 98(2):123–142, 1987.

[20] L. Modica. Gradient theory of phase transitions with boundary contact energy. In *Annales de l'IHP Analyse non linéaire*, volume 4, pages 487–512, 1987.

[21] A. Shapiro. On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 66(3):477–487, 1990.

[22] B. Stinner, B. Nestler, and H. Garcke. A diffuse interface model for alloys with multiple components and phases. *SIAM Journal on Applied Mathematics*, 64(3):775–799, 2004.

[23] M. Ulbrich. Semismooth Newton methods for operator equations in function spaces. *SIAM Journal on Optimization*, 13(3):805–841, 2002.

[24] B. Van Wachem and A.-E. Almstedt. Methods for multiphase computational fluid dynamics. *Chemical Engineering Journal*, 96(1):81–98, 2003.