FULLY DISCRETIZED SOBOLEV GRADIENT FLOW FOR THE GROSS-PITAEVSKII EIGENVALUE PROBLEM

ZIANG CHEN, JIANFENG LU, YULONG LU, AND XIANGXIONG ZHANG

ABSTRACT. This paper studies the numerical approximation of the ground state of the Gross-Pitaevskii (GP) eigenvalue problem with a fully discretized Sobolev gradient flow induced by the H^1 norm. For the spatial discretization, we consider the finite element method with quadrature using P^k basis on a simplicial mesh and Q^k basis on a rectangular mesh. We prove the global convergence to a critical point of the discrete GP energy, and establish a local exponential convergence to the ground state under the assumption that the linearized discrete Schrödinger operator has a positive spectral gap. We also show that for the P^1 finite element discretization with quadrature on an unstructured shape regular simplicial mesh, the eigengap satisfies a mesh-independent lower bound, which implies a mesh-independent local convergence rate for the proposed discrete gradient flow. Numerical experiments with discretization by high-order Q^k spectral element methods in two and three dimensions are provided to validate the efficiency of the proposed method.

1. INTRODUCTION

1.1. **The Gross-Pitaevskii eigenvalue problem.** A standard mathematical model of the equilibrium states in Bose–Einstein condensation (BEC) [11, 21, 27, 46] is through the minimization of the Gross-Pitaevskii energy. For N identical bosons, with an scattering length a and an external potential $V(\boldsymbol{x})$, the Gross-Pitaevskii (GP) energy functional is defined as

$$\mathcal{E}^{\mathrm{GP}}(\phi) = \int_{\mathbb{R}^3} \left(|\nabla\phi(\boldsymbol{x})|^2 + V(\boldsymbol{x})|\phi(\boldsymbol{x})|^2 + 4\pi a|\phi(\boldsymbol{x})|^4 \right) \mathrm{d}\boldsymbol{x},$$

and the GP energy, denoted by $E^{\mathrm{GP}}(N, a)$, is defined as the infimum of $\mathcal{E}^{\mathrm{GP}}$ under normalization $\int_{\mathbb{R}^3} |\phi(\boldsymbol{x})|^2 \mathrm{d}\boldsymbol{x} = N$. It has been used for finding the ground state energy per unit volume of a dilute, thermodynamically infinite, homogeneous gas. In some typical experiments the value of a is about 10^{-3}, while N varies from 10^3 to 10^7.

Computations are usually done on a finite domain and it is usually acceptable to make the assumption that we can approximate wave function of interest by compact support due to the fast decay rate at infinity [29, 38]. Let $\mathcal{E}_L^{\mathrm{GP}}$ denote the energy functional defined on a box domain $\Omega = [-L, L]^d$ and $E_L^{GP}(N, a)$ denote the corresponding GP energy, then $\lim_{L \to \infty} E_L^{GP}(N, a) = E^{GP}(N, a)$, see [44]. Since the GP energy satisfies the scaling relation $E^{GP}(N, a) = N E^{GP}(1, Na)$, we consider the following simplified rescaled model for finding ground state: minimizing the energy functional

$$E(\phi) = \frac{1}{2} \int_\Omega \left(|\nabla \phi(\boldsymbol{x})|^2 + V(\boldsymbol{x}) |\phi(\boldsymbol{x})|^2 \right) \mathrm{d}\boldsymbol{x} + \frac{\beta}{4} \int_\Omega |\phi(\boldsymbol{x})|^4 \mathrm{d}\boldsymbol{x}, \quad \Omega = [-L, L]^d,$$

over the constraint set $\left\{ \phi \in H_0^1(\Omega) : \int_\Omega |\phi(\boldsymbol{x})|^2 \mathrm{d}\boldsymbol{x} = 1 \right\}$, where $d = 1, 2, 3$, $V(\boldsymbol{x}) \geq 0$ and $\beta > 0$. While it is also of interest to extend the problem to $\beta < 0$, see e.g., [3], we restrict to the case $\beta > 0$ in this work.

The existence, uniqueness, and regularity of the GP ground states are well understood, see e.g., [44]. For $\beta > 0$, $E(\phi)$ has the unique positive ground state $\phi(\boldsymbol{x}) > 0$, which is also the eigenfunction to the nonlinear eigenvalue problem
(1.1)

$$-\Delta u(\boldsymbol{x}) + V(\boldsymbol{x}) u(\boldsymbol{x}) + \beta |u(\boldsymbol{x})|^2 u(\boldsymbol{x}) = \lambda u(\boldsymbol{x}), \quad \int_\Omega |u(\boldsymbol{x})|^2 \mathrm{d}\boldsymbol{x} = 1, u(\boldsymbol{x})|_{\partial \Omega} = 0.$$

Notice that (1.1) should be understood in the sense of distribution, i.e., the variational form of (1.1) is to seek $\lambda \in \mathbb{R}$ and $u \in H_0^1(\Omega)$ satisfying

(1.2) $$(\nabla u, \nabla v) + (Vu, v) + \beta(|u|^2 u, v) = \lambda(u, v), \quad \forall v \in H_0^1(\Omega),$$

where $(u, v) = \int_\Omega u(\boldsymbol{x}) v(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$. Let u^* be the ground state to $E(\cdot)$, then by setting $u = v = u^*$ in (1.2), the corresponding eigenvalue should satisfy

$$\lambda^* = 2E(u^*) + \frac{\beta}{2} \bar{\rho}, \quad \bar{\rho} = \int_\Omega |\phi(\boldsymbol{x})|^4 \mathrm{d}\boldsymbol{x}.$$

Since the ground state u^* remains unchanged under a constant shift of the potential, without loss of generality, we may assume $V(\boldsymbol{x}) \geq c > 0$ for some $c > 0$.

1.2. **Related work.** The study of numerical solutions to the Gross-Pitaevskii problem (1.2) has a long history. Self-consistent field iteration (SCF) [14, 15, 24, 50] is one of the most popular iterative techniques for a nonlinear eigenvalue problem, which involves a linearized eigenvalue problem during each iteration. For the problem (1.1), SCF may diverge unless a good initial guess is provided.

Another category of popular methods takes an optimization perspective of the energy functional. They can be viewed as discrete-in-time gradient flows (i.e., gradient descent) of the energy functional linked to (1.2). Earlier works in this category are based on an implicit Euler discretization of the L^2-gradient flow [8–10]. More recently, several alternative gradient flows have been proposed by modifying the underlying metric, including the projected Sobolev gradient flow [16, 22, 23, 31, 33, 36, 53] and the J-method [2, 35]. Projected Sobolev gradient flow is based on first computing the Sobolev gradients, which are the Riesz representation of the Fréchet derivative of the GP energy functional within an appropriate Hilbert space (e.g., $H^1(\Omega)$), and then projecting the gradients to the tangent space of the Riemannian manifold defined by the normalization constraint. Despite the empirical success

of projected Sobolev gradient flows for solving the GP eigenvalue problem, their convergence analysis is still underdeveloped. Our work fits in this line of research.

For the existing convergence results for gradient-flow-based methods: The work [36] established the global exponential convergence of the continuous-in-time projected H^1-gradient flow to a critical point of E. The work [33] obtained a global exponential convergence of a continuous projected Sobolev flow with an alternative metric to the ground state and also proved the global convergence (without a rate) of its forward Euler discretization. A more recent work [53] established a local exponential convergence of the discrete-in-time flow of [33] under the assumption that the discrete iterates are uniformly bounded. A more explicit local convergence rate depending on the eigengap of the linearized problem at the ground state is obtained in [32]. In our previous work [16], we improved the analysis of the global convergence and local rate of convergence of discrete-in-time projected Sobolev gradient flows with several common choices of inner products in H^1-space.

In addition to the time-discretization of the projected Sobolev gradient flows, their spatial discretization [8] is of course necessary for the practical implementation of the schemes. However, most of the prior theoretical work on projected Sobolev flows for the GP eigenvalue problem does not consider spatial discretization and it remains open how to extend the convergence analysis to the fully discretized setting. The convergence of the numerical solution using finite element method has been first established in [13], and there is also some recent progress on estimating the discretization error for energy, eigenvalue, and eigenfunction in the setting of mixed finite element method [28], but the convergence of the fully discretized gradient flows has not been analyzed before the initial submission of this work. After the initial submission of this work, some very recent progress was reported in [30], which extended the results in [33] to the fully discretized A_u Sobolev gradient descent with the monotone P^1 finite element method. Compared to [30], we consider the fully discretized H^1 Sobolev gradient descent and we also prove locally exponential convergence rate that is not covered in [30].

Let us also mention some works on numerical analysis for general nonlinear eigenvalue problems, where $\frac{\beta}{4} \int_\Omega |\phi(\boldsymbol{x})|^4 \mathrm{d}\boldsymbol{x}$ in the energy functional of the GP eigenvalue problem is generalized to $\frac{1}{2} \int_\Omega F(|\phi(\boldsymbol{x})|^2) \mathrm{d}\boldsymbol{x}$; we refer interested readers to [12, 13, 26].

1.3. **Contribution of the present work.** We summarize our major contribution as follows.

- We propose a fully discretized Sobolev gradient descent for approximating the ground state of the GP energy, which can be viewed as a Riemannian gradient descent method on the sphere under a metric induced by a modified H^1-norm.
- We prove the global convergence of the fully discretized Sobolev gradient descent with respect to the modified H^1 metric to a critical point of the discrete GP energy and a local convergence to the ground state with an exponential rate. See Corollary 5.3 and Theorem 5.10. We prove the convergence of the ground eigenpair of the discrete GP energy to those of the continuous counterpart as well as a positive discrete eigengap as the mesh size diminishes for P^1 finite element method with quadrature on unstructured shape regular simplicial meshes under the classical mesh constraint

for monotonicity, which includes the second-order finite difference scheme; see Theorem E.6 and Theorem E.8.
- We provide numerical experiments with Q^k spectral element method as spatial discretization to verify the accuracy and efficiency of the proposed approach for solving GP problems in both two and three dimensions. Due to the fact that only Laplacian needs to be inverted in the algorithm, the scheme on a structured mesh can be easily implemented and efficiently accelerated on modern GPUs.

1.4. **Organization of the rest of the paper.** As preliminaries, we first discuss the spatial discretization of the GP eigenvalue problem based on finite element method in Section 2, then review some useful properties of the discrete energy in Section 3. In particular, the second-order finite element method on a uniform gives the most popular second-order finite difference scheme. The fully discretized Sobolev gradient descent methods are given in Section 4. We present the two main convergence results in Section 5, with numerical experiments given in Section 6. Further preliminary results and proof details can be found in the Appendix. Concluding remarks are given in Section 7.

2. The classical finite element method with quadrature

We consider the classical continuous finite element method with quadrature using P^k basis on a simplicial mesh or Q^k basis on a rectangular mesh. This section briefly reviews its definition. If the finite element method is defined on a uniform structured mesh, it is well known that both the P^1 scheme and the Q^1 scheme are equivalent to the second-order finite difference scheme.

2.1. **Finite element Galerkin method.** We first consider a uniform rectangular mesh Ω_h for the rectangular domain Ω. For any rectangle e in the mesh Ω_h, let Q^k be the space of tensor product polynomials of degree k:

$$Q^k(e) = \left\{ p(\boldsymbol{x}) = \sum_{i_1, i_2, \ldots, i_d = 0}^{k} p_{i_1 i_2 \cdots i_d} x_1^{i_1} x_2^{i_2} \cdots x_d^{i_d}, \ \boldsymbol{x} = (x_1, x_2, \ldots, x_d) \in e \right\}.$$

Let $V_0^h \subset H_0^1(\Omega)$ be continuous piecewise Q^k polynomial space with zero boundary:

$$V_0^h = \{ v_h(\boldsymbol{x}) \in C(\Omega) : v_h(\boldsymbol{x})|_{\partial\Omega} = 0, \ v_h\big|_e \in Q^k(e), \ \forall e \in \Omega_h \} \subset H_0^1(\Omega).$$

We also consider an unstructured simplicial mesh Ω_h with e denoting a simplex in Ω_h, e.g., a triangular mesh in two dimensions with e denoting a triangle. Let P^k be the space of polynomials of degree k:

$$P^k(e) = \left\{ p(\boldsymbol{x}) = \sum_{i_1 + i_2 + \cdots + i_d \leq k} p_{i_1 i_2 \cdots i_d} x_1^{i_1} x_2^{i_2} \cdots x_d^{i_d}, \ \boldsymbol{x} = (x_1, x_2, \ldots, x_d) \in e \right\},$$

and $V_0^h \subset H_0^1(\Omega)$ be continuous piecewise P^k polynomial space with zero boundary:

$$V_0^h = \{ v_h(\boldsymbol{x}) \in C(\Omega) : v_h(\boldsymbol{x})|_{\partial\Omega} = 0, \ v_h\big|_e \in P^k(e), \ \forall e \in \Omega_h \} \subset H_0^1(\Omega).$$

The finite element Galerkin method for (1.1) is to seek $\lambda_h \in \mathbb{R}$ and $u_h \in V_0^h$ satisfying

$$(2.1) \qquad (\nabla u_h, \nabla v_h) + (V u_h, v_h) + \beta(|u_h|^2 u_h, v_h) = \lambda_h(u_h, v_h), \quad \forall v_h \in V_0^h.$$

The corresponding discrete energy can be given as

$$E(u_h) = \frac{1}{2}(\nabla u_h, \nabla u_h) + \frac{1}{2}(V u_h, u_h) + \frac{\beta}{4}(u_h^2, u_h^2).$$

The convergence of the finite element method for the nonlinear eigenvalue problem (1.1) was discussed in [13]. The standard *a priori* error estimates for a linear eigenvalue problem, e.g., $\beta = 0$ in (2.1) is $2k$-order for eigenvalues and k-order for eigenvector in H^1-norm, under suitable regularity assumptions. See [4–6,17,37] for discussions on the rate of convergence of numerical schemes for eigenvalue problems.

2.2. Finite element method with quadrature.
In practice, one often uses quadrature for integrals to implement the finite element method. The Q^k spectral element method is to replace all integrals in (2.1) by $(k + 1)$-point Gauss-Lobatto quadrature in each dimension. Standard *a priori* finite element method error estimates still hold, see [18] and references therein. For the P^k finite element method, suitable quadrature on a simplex can be used, e.g., the simplest quadrature using the average of values at vertices can be used for the P^1 finite element method. Let $\langle \cdot, \cdot \rangle$ denote that integrals are replaced by quadrature, then the method is to find $u_h \in V^h$ satisfying

$$(2.2) \qquad \langle \nabla u_h, \nabla v_h \rangle + \langle V u_h, v_h \rangle + \beta \langle |u_h|^2 u_h, v_h \rangle = \lambda_h \langle u_h, v_h \rangle, \quad \forall v_h \in V_0^h.$$

The corresponding discrete energy is given as

$$(2.3) \qquad E_h(u_h) = \frac{1}{2}\langle \nabla u_h, \nabla u_h \rangle + \frac{1}{2}\langle V u_h, u_h \rangle + \frac{\beta}{4}\langle u_h^2, u_h^2 \rangle.$$

2.3. The matrix-vector form.
For either Q^k or P^k finite element method, we will use a quadrature rule such that a matrix-vector form of the scheme can be easily written.

We first describe the Q^k finite element method on a uniform rectangular mesh. Assume that Ω_h consists of uniform N_c^d cubic cells for the cubic domain $\Omega = [-L, L]^d$. Then there are in total $(N_c k + 1)^d$ Gauss-Lobatto points. Any Q^k polynomial on a cubic element e can be represented as a Lagrangian interpolation polynomial at $(k + 1)^d$ Gauss-Lobatto points, thus the Q^k spectral element method (2.2) also becomes a finite difference scheme on all Gauss-Lobatto nodes. For Q^1 and Q^2 bases, all the Gauss-Lobatto points form a uniform grid. For $k \geq 3$, the Gauss-Lobatto points are not uniform in each element. For homogeneous Dirichlet boundary condition, the boundary points are not unknown. Thus the total number of unknowns is the interior grid points with the number $N = n^d$ where $n = N_c k - 1$. To derive an equivalent matrix form of the scheme (2.2), let $\phi_i(\boldsymbol{x}) \in V_0^h$ $(i = 1, \cdots, N)$ be the continuous piecewise Q^k Lagrangian basis at all Gauss-Lobatto points $\boldsymbol{x}_i$ $(i = 1, \cdots, N)$ in the interior of Ω_h. For any piecewise polynomial $u_h(\boldsymbol{x}) \in V_0^h$, let $u_i = u_h(\boldsymbol{x}_i)$. Then $u_h(\boldsymbol{x}) = \sum_{i=1}^N u_i \phi_i(\boldsymbol{x})$. Let $\mathbf{u} = \begin{bmatrix} u_1 & \cdots & u_N \end{bmatrix}^\top$ and w_i be the quadrature weight at $\boldsymbol{x}_i$.

Next we consider the P^k finite element method on a simplicial mesh. The number of degree of freedoms of P^k polynomial on a d-dimensional simplex is $\frac{(k+d)!}{k!d!}$. Consider a $(k + 1)$-th order accurate quadrature rule on a simplex using $\frac{(k+d)!}{k!d!}$ quadrature points, e.g., the quadrature rule using 3 vertices for P^1 on a triangle, the quadrature rule using 6 vertices for P^1 on a tetrahedron and the quadrature rule using 3 vertices and 3 edge centers for P^2 on a triangle. Assume there are

in total N quadrature points which lie in the interior of the domain Ω_h. Let $\boldsymbol{x}_i$ $(i = 1, \cdots, N)$ be all interior quadrature points with w_i being the quadrature weight at $\boldsymbol{x}_i$. Let $\phi_i(\boldsymbol{x}) \in V_0^h$ $(i = 1, \cdots, N)$ be the continuous piecewise P^k Lagrangian basis at all interior quadrature points $\boldsymbol{x}_i$ $(i = 1, \cdots, N)$. For any piecewise polynomial $u_h(\boldsymbol{x}) \in V_0^h$, let $u_i = u_h(\boldsymbol{x}_i)$. Then $u_h(\boldsymbol{x}) = \sum_{i=1}^{N} u_i \phi_i(\boldsymbol{x})$. Let $\mathbf{u} = \begin{bmatrix} u_1 & \cdots & u_N \end{bmatrix}^\top$.

With the notation above, we have

$$(2.4) \qquad \langle V u_h, v_h \rangle = \sum_{i=1}^{N} w_i V_i u_i v_i = \mathbf{v}^\top \mathbb{M} \mathbb{V} \mathbf{u},$$

where $\mathbb{M} = \mathrm{diag}\{w_1, \cdots, w_N\}$ and $\mathbb{V} = \mathrm{diag}\{V_1, \cdots, V_N\}$ are diagonal matrices and $V_i = V(\boldsymbol{x}_i)$. We also have

$$(2.5) \qquad \langle \nabla u_h, \nabla v_h \rangle = \mathbf{v}^\top \mathbb{S} \mathbf{u},$$

where $\mathbb{S}$ is the stiffness matrix given by $\mathbb{S}_{ij} = \langle \nabla \phi_i, \nabla \phi_j \rangle$.

Using (2.4) and (2.5), the matrix form of (2.2) is to find $\mathbf{u} \in \mathbb{R}^N$ satisfying

$$\mathbf{v}^\top \mathbb{S} \mathbf{u} + \mathbf{v}^\top \mathbb{M} \mathbb{V} \mathbf{u} + \beta \mathbf{v}^\top \mathbb{M} \mathbf{u}^3 = \lambda_h \mathbf{v}^\top \mathbb{M} \mathbf{u}, \quad \forall \mathbf{v} \in \mathbb{R}^N,$$

or equivalently

$$(2.6) \qquad \mathbb{S} \mathbf{u} + \mathbb{M} \mathbb{V} \mathbf{u} + \beta \mathbb{M} \mathbf{u}^3 = \lambda_h \mathbb{M} \mathbf{u},$$

where $\mathbf{u}^3 = \begin{bmatrix} u_1^3 & \cdots & u_N^3 \end{bmatrix}^\top$. Let $\Delta_h = -\mathbb{M}^{-1} \mathbb{S}$, then (2.6) can also be written as

$$(2.7) \qquad -\Delta_h \mathbf{u} + \mathbb{V} \mathbf{u} + \beta \mathbf{u}^3 = \lambda_h \mathbf{u}.$$

In the formulation (2.6), all the matrices are symmetric positive definite, but the discrete Laplacian $\Delta_h = -\mathbb{M}^{-1} \mathbb{S}$ is in general not symmetric in (2.7) except the special case of second-order finite difference, which however does not affect numerical implementations since the symmetric form (2.6) should be implemented instead of the form (2.7).

2.4. **The discrete energy and discrete L^2 norm.** Using the same notation, the discrete energy (2.3) can be written as

$$(2.8) \qquad E_h(u_h) = \frac{1}{2} \mathbf{u}^\top \mathbb{S} \mathbf{u} + \frac{1}{2} \mathbf{u}^\top \mathbb{M} \mathbb{V} \mathbf{u} + \frac{\beta}{4} (\mathbf{u}^2)^\top \mathbb{M} \mathbf{u}^2.$$

Introduce the interpolation operator

$$(2.9) \qquad \Pi : \mathbb{R}^N \longrightarrow V_0^h, \quad \mathbf{v} \longmapsto \sum_{i=1}^{N} v_i \phi_i(\boldsymbol{x}).$$

The discrete integration by parts is ensured in the following sense:

$$(2.10) \qquad \langle \nabla u_h, \nabla v_h \rangle = \mathbf{v}^\top \mathbb{S} \mathbf{u} = \mathbf{v}^\top \mathbb{M} (-\Delta_h) \mathbf{u} = \langle -\Pi[\Delta_h \mathbf{u}], v_h \rangle.$$

For two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$, we define the discrete L^2 inner product $\langle \mathbf{u}, \mathbf{v} \rangle_h$ by setting

$$(2.11) \qquad \langle \mathbf{u}, \mathbf{v} \rangle_h := \mathbf{u}^\top \mathbb{M} \mathbf{v}.$$

Thus the discrete energy (2.8) and (2.3) can also be written in matrix form

$$(2.12) \qquad E_h(u_h) = \mathbf{E}_h(\mathbf{u}) = \frac{1}{2} \langle -\Delta_h \mathbf{u}, \mathbf{u} \rangle_h + \frac{1}{2} \langle \mathbb{V} \mathbf{u}, \mathbf{u} \rangle_h + \frac{\beta}{4} \langle \mathbf{u}^2, \mathbf{u}^2 \rangle_h,$$

with normalization constraint $\langle \mathbf{u}, \mathbf{u} \rangle_h = 1$.

3. Properties of the discrete energy

We discuss some properties of the discrete energy $E_h(u_h) = \mathbf{E}_h(\mathbf{u})$ in this section.

3.1. The monotonicity of the discrete Laplacian.

A matrix $A \in \mathbb{R}^{n \times n}$ is called *monotone* if its inverse has nonnegative entries $A^{-1} \geq 0$. At a fixed vector $\mathbf{u}$, the linearized operator for (2.7) is given by

$$(3.1) \qquad A_{\mathbf{u}} = -\Delta_h + \mathbb{V} + \beta \operatorname{diag}(\mathbf{u}^2).$$

The matrix $A_{\mathbf{u}}$ is *irreducible*, which can be easily verified by the graph that the discrete Laplacian represents, see [41]. The definition of irreducible matrices is given in Appendix B.2. By the Perron Frobenius Theorem in Appendix B.2, if $A_{\mathbf{u}}$ is also monotone, then its smallest eigenvalue has multiplicity one, with a unique unit positive eigenvector. For the Q^1 finite element scheme on a uniform rectangular mesh (or equivalently the second-order finite difference scheme) or the P^1 finite element method on a simplical mesh under suitable constraints is used, it is straightforward to verify that $A_{\mathbf{u}}$ satisfies Theorem B.1, implying that $A_{\mathbf{u}}$ is an M-matrix thus monotone, which is a well-known result in the literature; see Appendix B. The explicit expressions of the second-order finite difference and P^1 finite element method on an unstructured mesh are given in Appendix A.

For a simplex T in a simplicial mesh $\Omega_h \subset \mathbb{R}^d$ of dimension d, let κ_E^T be the $(d-2)$-dimensional simplex opposite to the edge E in the simplex T and θ_E^T be the angle between the two faces containing the edge E in the simplex T. By [52, Lemma 2.1], the simplicial mesh constraint for monotonicity is

$$(3.2) \qquad \sum_{T \supset E} \frac{1}{d(d-1)} |\kappa_E^T| \cot \theta_E^T \geq 0,$$

where $T \supset E$ means summation over all simplexes T containing the edge E. Such a constraint reduces to a Delaunay triangular mesh in two dimensions (see Appendix A.2), which is more general and more practical than a non-obtuse triangulation. We summarize the results in Theorem 3.1:

Theorem 3.1. *For the P^1 finite element scheme with quadrature on a simplicial mesh satisfying (3.2), which includes the classical second-order finite difference scheme, $A_{\mathbf{u}} = -\Delta_h + \mathbb{V} + \beta \operatorname{diag}(\mathbf{u}^2)$ is an M-matrix thus monotone. As a result, it has a unique positive unit eigenvector and the corresponding eigenvalue is simple and the smallest eigenvalue of $A_{\mathbf{u}}$.*

For the high-order accurate discrete Laplacian, the matrix $-\Delta_h + \mathbb{V}$ is no longer an M-matrix. It is proven in [41] that the fourth-order accurate Laplacian of Q^2 scheme in two dimensions are products of M-matrices and thus still monotone under certain mesh size constraints. It is possible to prove similar results for the three-dimensional case following the same arguments in [41]. Extensions to quasi-uniform meshes are given in [19]. It is also possible to extend the monotonicity to Q^3 element [20]. All these monotonicity results for high-order schemes hold under mesh size constraints, which makes further discussion of global convergence much more complicated. Thus we only discuss the global convergence for the second-order scheme.

3.2. **Ground state of the discrete energy.** In this subsection, we only focus on P^1 finite element scheme on a simplicial mesh satisfying (3.2), which includes the second-order finite difference scheme. In general, it is difficult to extend all results about the discrete energy (2.12) to high-order schemes.

Theorem 3.2. *For the P^1 finite element method with quadrature weights w_i on a simplicial mesh satisfying (3.2), for any $v_h \in V_0^h$ satisfying $\sum_{i=1}^N w_i v_i = 1$ and $v_i \geq 0$, $\forall i$, where $v_i = v_h(\mathbf{x}_i)$, $\mathbf{E}_h(\sqrt{\mathbf{v}})$ is strongly convex w.r.t. the vector $\mathbf{v} = \begin{bmatrix} v_1 & \cdots & v_N \end{bmatrix}^\top$.*

Proof. Let $u_h \in V_0^h$ satisfy $u_h(\mathbf{x}_i) = \sqrt{v_i} = \sqrt{v_h(\mathbf{x}_i)}$. By (2.3), we have

$$\mathbf{E}_h(\sqrt{\mathbf{v}}) = \frac{1}{2}\langle \nabla u_h, \nabla u_h \rangle + \frac{1}{2}\langle V u_h, u_h \rangle + \frac{\beta}{4}\langle u_h^2, u_h^2 \rangle.$$

The quadrature yields that $\frac{\beta}{4}\langle u_h^2, u_h^2 \rangle = \frac{\beta}{4}\sum_i w_i v_i^2$ is quadratic and strongly convex in $\mathbf{v}$ and that $\langle V u_h, u_h \rangle = \sum_i w_i V_i v_i$ is linear in $\mathbf{v}$. By (A.2), we have

$$\langle \nabla u_h, \nabla u_h \rangle = \sum_E \left(\sum_{T \supset E} \frac{1}{d(d-1)} |\kappa_E^T| \cot \theta_E^T \right) |\sqrt{v_i} - \sqrt{v_j}|^2,$$

the edge E connects vertices $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

With (3.2), the convexity of the term $\langle \nabla u_h, \nabla u_h \rangle$ is induced by the convexity of the bivariate function $f(x,y) = |\sqrt{x} - \sqrt{y}|^2$, which is easy to verify. $\square$

Theorem 3.3. *For the P^1 finite element method with quadrature on a simplicial mesh satisfying (3.2), $\forall u_h \in V_0^h$, $\mathbf{E}_h(\mathbf{u}) \geq \mathbf{E}_h(|\mathbf{u}|)$.*

Proof. It suffices to verify that $\langle \nabla u_h, \nabla u_h \rangle \geq \langle \nabla |u_h|, \nabla |u_h| \rangle$. By (A.2), we have

$$\langle \nabla u_h, \nabla u_h \rangle = \sum_E \left(\sum_{T \supset E} \frac{1}{d(d-1)} |\kappa_E^T| \cot \theta_E^T \right) |u_i - u_j|^2,$$

the edge E connects vertices $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

With (3.2), it suffices to verify $(x-y)^2 \geq (|x| - |y|)^2$, which is trivial. $\square$

Theorem 3.4. *For the P^1 finite element method with quadrature on a simplicial mesh satisfying (3.2), the discrete energy $E_h(u_h)$ under the constraint $\langle u_h, u_h \rangle = 1$ has a unique and positive minimizer u_h^*. Let $\mathbf{u}^*$ be the vector representing its point values $u_h^*(\boldsymbol{x}_i)$, then $\mathbf{u}^*$ solves (2.7), and $\mathbf{u}^*$ is the eigenvector associated to the smallest eigenvalue of the linear operator $A_{\mathbf{u}^*} = -\Delta_h + \mathbb{V} + \beta \operatorname{diag}(\mathbf{u}^*)^2$.*

Proof. Strong convexity over a convex constraint in Theorem 3.2 gives the existence and uniqueness of the minimizer u_h^*. Theorem 3.3 implies that $u_h^*(\boldsymbol{x}_i) \geq 0$. For minimizing $E_h(u_h)$ with $\langle u_h, u_h \rangle = 1$, or equivalently minimizing $\mathbf{E}_h(\mathbf{u})$ with $\mathbf{u}^T \mathbb{M} \mathbf{u} = 1$, the Lagrangian for the constrained minimization is given by $L(u_h, \lambda_h) = E_h(u_h) - \lambda_h(\langle u_h, u_h \rangle - 1)$. The minimizer must satisfy the critical point equation $\frac{\delta L}{\delta u_h} = 0$, thus u_h^* satisfies (2.2), or equivalently, $\mathbf{u}$ satisfies (2.7). By Theorem 3.1, the matrix $A_{\mathbf{u}} = -\Delta_h + \mathbb{V} + \beta \operatorname{diag}(\mathbf{u}^2)$ is monotone. By Perron Frobenius Theorem (Theorem B.3), such a monotone matrix $A_{\mathbf{u}^*}$ has a unique positive unit eigenvector associated with its smallest eigenvalue. Since $\mathbf{u}^* \geq 0$, it is the unique unit eigenvector to the smallest eigenvalue of the linear operator $A_{\mathbf{u}^*}$. $\square$

Remark 3.5. Neither Theorem 3.2 nor Theorem 3.3 can be extended to Q^2 finite element method. Nonetheless, the monotonicity of the fourth-order scheme may hold if the mesh size is very small, which ensures that the ground state of the non-linear eigenvalue problem gives rise to a positive eigenvector by Perron-Frobenius Theorem.

4. FULLY DISCRETIZED SOBOLEV GRADIENT DESCENT AND MODIFIED H^1 SCHEME

We define the schemes for minimizing the discrete energy (2.8) or (2.12) associated with the P^k or Q^k finite element method with suitable quadrature, under the normalization constraint:

$$(4.1) \qquad \mathcal{M} = \left\{ u_h \in V_0^h : \langle u_h, u_h \rangle = 1 \right\} = \left\{ \mathbf{u} \in \mathbb{R}^N : \langle \mathbf{u}, \mathbf{u} \rangle_h = \mathbf{u}^\top \mathbb{M} \mathbf{u} = 1 \right\},$$

where $u_h = \Pi(\mathbf{u})$ with Π being the interpolation operator defined in (2.9). The tangent space of the manifold $\mathcal{M}$ at u_h or $\mathbf{u}$ is

$$\mathcal{T}_{u_h} \mathcal{M} = \mathcal{T}_{\mathbf{u}} \mathcal{M} = \left\{ v_h \in V_0^h : \langle u_h, v_h \rangle = 0 \right\} = \left\{ \mathbf{v} \in \mathbb{R}^N : \langle \mathbf{u}, \mathbf{v} \rangle_h = 0 \right\}.$$

4.1. **The Sobolev gradient descent.** The gradient of the energy $E_h(u_h)$, say $\nabla E_h(u_h) \in V_0^h$, should be understood in the sense of the Fréchet derivative in the space V_0^h, and can be computed by

$$\begin{aligned}
\langle \nabla E_h(u_h), v_h \rangle &= \lim_{t \to 0} \frac{E_h(u_h + t v_h) - E_h(u_h)}{t} \\
&= \mathbf{v}^\top \mathbb{S} \mathbf{u}^\top + \mathbf{v}^\top \mathbb{M} \mathbb{V} \mathbf{u} + \beta \mathbf{v}^\top \mathbb{M} \mathbf{u}^3 \\
&= \langle \nabla u_h, \nabla v_h \rangle + \langle V u_h, v_h \rangle + \beta \langle \Pi(\mathbf{u}^3), v_h \rangle,
\end{aligned}$$

for all $v_h \in V_0^h$. With the discrete integration by parts (2.10), we get

$$\nabla E_h(u_h) = -\Pi[\Delta_h \mathbf{u}] + \Pi(\mathbb{V} \mathbf{u}) + \beta \Pi(\mathbf{u}^3).$$

Similarly, since $\langle \mathbf{u}, \mathbf{v} \rangle_h = \mathbf{u}^\top \mathbb{M} \mathbf{v}$ as in (2.11), the gradient $\nabla \mathbf{E}_h(\mathbf{u}) \in \mathbb{R}^N$ is given by

$$\nabla \mathbf{E}_h(\mathbf{u}) = \mathbb{M}^{-1} (\mathbb{S} \mathbf{u} + \mathbb{M} \mathbb{V} \mathbf{u} + \beta \mathbb{M} \mathbf{u}^3) = (-\Delta_h + \mathbb{V}) \mathbf{u} + \beta \mathbf{u}^3 = A_{\mathbf{u}} \mathbf{u},$$

where $A_{\mathbf{u}}$ is defined in (3.1). Thus the two Fréchet derivatives $\nabla E_h(u_h)$ and $\nabla \mathbf{E}_h(\mathbf{u})$ are also identical in the sense that $\nabla E_h(u_h) = \Pi[\nabla \mathbf{E}_h(\mathbf{u})]$.

Given any inner product $\langle \cdot, \cdot \rangle_X$ on $\mathbb{R}^N$, one can equip the manifold $\mathcal{M} \subset \mathbb{R}^N$ defined in (4.1) with an Riemannian metric $g(\mathbf{u}, \mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle_X$. Let $\mathbb{G}_X \in \mathbb{R}^{N \times N}$ be the positive definite matrix satisfying

$$\langle \mathbf{u}, \mathbb{G}_X \mathbf{v} \rangle_X = \langle \mathbf{u}, \mathbf{v} \rangle_h, \quad \forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^N.$$

The *Riemannian gradient* of $\mathbf{E}_h(\mathbf{u})$ at $\mathbf{u} \in \mathcal{M}$ is defined as $\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}) \in \mathcal{T}_{\mathbf{u}} \mathcal{M}$ satisfying

$$g\left(\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}), \mathbf{v} \right) = \langle \nabla \mathbf{E}_h(\mathbf{u}), \mathbf{v} \rangle_h, \quad \forall \mathbf{v} \in \mathcal{T}_{\mathbf{u}} \mathcal{M}.$$

Following a similar derivation of Fréchet derivatives as before, the gradient of the discrete energy with respect to the inner product $\langle \cdot, \cdot \rangle_X$ can be computed as

$$\nabla_X \mathbf{E}_h(\mathbf{u}) = \mathbb{G}_X A_{\mathbf{u}} \mathbf{u}.$$

For any $\mathbf{w} \in \mathbb{R}^N$, the projection of $\mathbf{w}$ onto $\mathcal{T}_{\mathbf{u}} \mathcal{M}$ with respect to $\langle \cdot, \cdot \rangle_X$ is given by

$$\mathcal{P}_{\mathcal{T}_{\mathbf{u}} \mathcal{M}, X}(\mathbf{w}) = \mathbf{w} - \frac{\langle \mathbf{u}, \mathbf{w} \rangle_h}{\langle \mathbf{u}, \mathbb{G}_X \mathbf{u} \rangle_h} \mathbb{G}_X \mathbf{u}.$$

Therefore, the Riemannian gradient is

$$(4.2) \qquad \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}) = \mathcal{P}_{\mathcal{T}_{\mathbf{u}}\mathcal{M}, X} \left(\nabla_X \mathbf{E}_h(\mathbf{u}) \right) = \nabla_X \mathbf{E}_h(\mathbf{u}) - \frac{\langle \mathbf{u}, \nabla_X \mathbf{E}_h(\mathbf{u}) \rangle_h}{\langle \mathbb{G}_X \mathbf{u}, \mathbf{u} \rangle_h} \mathbb{G}_X \mathbf{u},$$

and the Riemannian gradient descent of minimizing $\mathbf{E}_h(\mathbf{u})$ over $\mathcal{M}$ with step size τ is

$$\mathbf{u}^{n+1} = R_h \left(\mathbf{u}^n - \tau \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n) \right), \quad R_h(\mathbf{u}) = \frac{\mathbf{u}}{\sqrt{\langle \mathbf{u}, \mathbf{u} \rangle_h}},$$

where $R_h(\mathbf{u})$ is the retraction operator approximating the exponential map [1].

4.2. The modified H^1 scheme.

Different choices of the inner product $\langle \mathbf{u}, \mathbf{v} \rangle_X$ or the Riemannian metric g would lead to different schemes. In this work, we mainly focus on the modified H^1 scheme. In particular, for some constant $\alpha > 0$ the inner product $(\nabla u, \nabla v) + \alpha(u, v)$ for $H_0^1(\Omega)$ gives the following discrete inner product or metric

$$(4.3)$$
$$g(\mathbf{u}, \mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle_X = \langle \nabla u_h, \nabla v_h \rangle + \alpha \langle u_h, v_h \rangle = \langle \mathbf{u}, (-\Delta_h + \alpha \mathbb{I}) \mathbf{v} \rangle_h = \mathbf{u}^\top (\mathbb{S} + \alpha \mathbb{M}) \mathbf{v}.$$

This induces $\mathbb{G}_X = (-\Delta_h + \alpha \mathbb{I})^{-1}$ and

$$(4.4a) \qquad \nabla_X \mathbf{E}_h(\mathbf{u}) = \mathbb{G}_X A_{\mathbf{u}} \mathbf{u} = (-\Delta_h + \alpha \mathbb{I})^{-1}(-\Delta_h + \mathbb{V} + \beta \mathrm{diag}(\mathbf{u})^2) \mathbf{u}.$$

The corresponding Riemannian gradient is

$$(4.4b) \qquad \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}) = \nabla_X E_h(\mathbf{u}) - \frac{\langle \mathbf{u}, \nabla_X E_h(\mathbf{u}) \rangle_h}{\langle \mathbf{u}, (-\Delta_h + \alpha \mathbb{I})^{-1} \mathbf{u} \rangle_h} (-\Delta_h + \alpha \mathbb{I})^{-1} \mathbf{u},$$

and the Riemannian gradient descent method or the Sobolev gradient flow under the modified H^1-norm is hence given by

$$(4.4c) \qquad \mathbf{u}^{n+1} = R_h \left(\mathbf{u}^n - \tau \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n) \right).$$

If $\alpha = 0$, then (4.4) is the H^1 gradient flow algorithm in [33]. There are algorithms induced by other more complicated Riemannian metrics such as a_0-scheme with $g_u(w, z) = (\nabla w, \nabla z) + (w, Vz)$ and the a_u-scheme with $g_u(w, z) = (\nabla w, \nabla z) + (w, Vz) + \beta(w, u^2 z)$. We refer interested readers to [16, 33, 53].

5. Global and local convergence

This section proves the convergence of the modified H^1 scheme (4.4). The theories are inspired by our prior work [16] without spatial discretization and the main difficulty/novelty of this work is the analysis of the discrete schemes and discrete eigengap.

5.1. Energy decay and global convergence.

We define the discrete L^2 norm and the X-norm with a fixed parameter $\alpha > 0$ as follows:

$$\|\mathbf{u}\|_2 = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle_h} = \sqrt{\mathbf{u}^\top \mathbb{M} \mathbf{u}},$$

$$\|\mathbf{u}\|_X = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle_X} = \sqrt{\langle \mathbf{u}, \mathbb{G}_X^{-1} \mathbf{u} \rangle_h} = \sqrt{\mathbf{u}^\top (\mathbb{S} + \alpha \mathbb{M}) \mathbf{u}} = \sqrt{\langle \mathbf{u}, (-\Delta_h + \alpha \mathbb{I}) \mathbf{u} \rangle_h}.$$

Note that we have omitted the dependence of X-norm on α in the above. The main theorem in this subsection is stated as follows, which quantitatively characterizes the energy decay property of the Sobolev gradient flow under the modified H^1-norm (4.4).

Assumption 5.1. The potential energy V satisfies that $V_{\min} \leq V \leq V_{\max}$ for some constants $V_{\min}, V_{\max} \in (0, +\infty)$.

Theorem 5.2. *Suppose that Assumption* 5.1 *holds. Let* $\mathbf{u}^0 \in \mathcal{M} \subset \mathbb{R}^N$ *and let* $\{\mathbf{u}^n\}_{n=0}^{\infty}$ *be the sequence generated by* (4.4). *There exists constants* $C_u, C_g, C_\tau > 0$ *depending only on* $\|\mathbf{u}^0\|_X$ *and* $\Omega, V_{\min}, V_{\max}, \alpha, \beta, k$, *such that as long as* $0 < \tau_{\min} \leq \tau_n \leq \tau_{\max} \leq \min\{1, C_\tau\}$, $\forall n \geq 0$, *the followings holds for any* $n \geq 0$:

(i) $\|\mathbf{u}^n\|_X \leq C_u$.
(ii) $\left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X \leq \|\nabla_X \mathbf{E}_h(\mathbf{u}^n)\|_X \leq C_g$.
(iii) $\mathbf{E}_h(\mathbf{u}^n) - \mathbf{E}_h(\mathbf{u}^{n+1}) \geq C_d \left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X^2$, *where* $C_d = \frac{\tau_{\min}}{2}$.

A direct corollary is the global convergence to a critical point.

Corollary 5.3. *In the same setting as in Theorem* 5.2, *every limit point of* $\{\mathbf{u}^n\}_{n=0}^{\infty}$ *is a critical point.*

The rest of this subsection is for proving Theorem 5.2 and Corollary 5.3. We need a sequence of lemmas with the proofs of the first two lemmas being deferred to Appendix C.

Lemma 5.4. *For any* $\mathbf{u} \in \mathcal{M}$ *and* $\mathbf{v} \in \mathcal{T}_{\mathbf{u}}\mathcal{M}$, *it holds that*

$$(5.1) \qquad \|R_h(\mathbf{u} + \mathbf{v}) - (\mathbf{u} + \mathbf{v})\|_X \leq \frac{1}{2} \|\mathbf{v}\|_2^2 \|\mathbf{u} + \mathbf{v}\|_X.$$

Lemma 5.5. *There exists positive constants* C_1, C_2 *independent of the mesh size* h *such that for any* $\mathbf{u} \in \mathbb{R}^N$:

(i) $\|\mathbf{u}\|_2 \leq \frac{1}{\sqrt{\alpha}} \|\mathbf{u}\|_X$;
(ii) $\left\|(-\Delta_h + \alpha\mathbb{I})^{-1}\mathbf{u}\right\|_X \leq \frac{1}{\sqrt{\alpha}} \|\mathbf{u}\|_2$;
(iii) $\|\mathbf{u}^2\|_2 \leq C_1 \|\mathbf{u}\|_X^2$;
(iv) $\|\mathbf{u}^3\|_2 \leq C_2 \|\mathbf{u}\|_X^3$.

Lemma 5.6. *Let* C_2 *be the one in Lemma* 5.5 *and* $V_{\alpha,\max} = \|V - \alpha\|_{L^\infty(\Omega)}$. *Then it holds for any* $\mathbf{u} \in \mathcal{M}$ *that*

$$\langle \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}), \nabla_X \mathbf{E}_h(\mathbf{u}) \rangle_X = \|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u})\|_X^2,$$

$$(5.2) \qquad \left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u})\right\|_X \leq \|\nabla_X \mathbf{E}_h(\mathbf{u})\|_X \leq \|\mathbf{u}\|_X + \frac{V_{\alpha,\max}}{\alpha} \|\mathbf{u}\|_X + \frac{\beta C_2}{\sqrt{\alpha}} \|\mathbf{u}\|_X^3.$$

Proof. Since $\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}) \in \mathcal{T}_{\mathbf{u}}\mathcal{M}$ implies $\left\langle \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}), \mathbf{u} \right\rangle_h = 0$,

$$\left\langle \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}), \frac{\langle \mathbf{u}, \nabla_X \mathbf{E}_h(\mathbf{u}) \rangle}{\langle \mathbb{G}_X \mathbf{u}, \mathbf{u} \rangle} \mathbb{G}_X \mathbf{u} \right\rangle_X = \frac{\langle \mathbf{u}, \nabla_X \mathbf{E}_h(\mathbf{u}) \rangle_h}{\langle \mathbb{G}_X \mathbf{u}, \mathbf{u} \rangle_h} \left\langle \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}), \mathbf{u} \right\rangle_h = 0.$$

Due to $\nabla_X \mathbf{E}_h(\mathbf{u}) = \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}) + \frac{\langle \mathbf{u}, \nabla_X \mathbf{E}_h(\mathbf{u}) \rangle_h}{\langle \mathbb{G}_X \mathbf{u}, \mathbf{u} \rangle_h} \mathbb{G}_X \mathbf{u}$, the first equation holds, and

$$\|\nabla_X \mathbf{E}_h(\mathbf{u})\|_X^2 = \left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u})\right\|_X^2 + \left\|\frac{\langle \mathbf{u}, \nabla_X \mathbf{E}_h(\mathbf{u}) \rangle}{\langle \mathbb{G}_X \mathbf{u}, \mathbf{u} \rangle} \mathbb{G}_X \mathbf{u}\right\|_X^2$$

$$\Rightarrow \left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u})\right\|_X \leq \|\nabla_X \mathbf{E}_h(\mathbf{u})\|_X.$$

Let $\mathbb{V}_\alpha = \mathbb{V} - \alpha I_N$, then $-\Delta_h + \mathbb{V} = (-\Delta_h + \alpha\mathbb{I}) + \mathbb{V}_\alpha$. By (4.4a) and Lemma 5.5,

$$\|\nabla_X \mathbf{E}_h(\mathbf{u})\|_X \leq \|\mathbf{u}\|_X + \left\|(-\Delta_h + \alpha\mathbb{I})^{-1}(\mathbb{V}_\alpha \mathbf{u})\right\|_X + \beta \left\|(-\Delta_h + \alpha\mathbb{I})^{-1}(\mathbf{u}^3)\right\|_X$$

$$\leq \|\mathbf{u}\|_X + \frac{1}{\sqrt{\alpha}} \|\mathbb{V}_\alpha \mathbf{u}\|_2 + \frac{\beta C_2}{\sqrt{\alpha}} \|\mathbf{u}\|_X^3 \leq \|\mathbf{u}\|_X + \frac{V_{\alpha,\max}}{\alpha} \|\mathbf{u}\|_X + \frac{\beta C_2}{\sqrt{\alpha}} \|\mathbf{u}\|_X^3.$$

$\square$

Lemma 5.7. *Let C_1 be the constant in Lemma* 5.5. *For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$,*

$$|\mathbf{E}_h(\mathbf{u} + \mathbf{v}) - \mathbf{E}_h(\mathbf{u}) - \langle \nabla_X \mathbf{E}_h(\mathbf{u}), \mathbf{v} \rangle_X|$$
$$\leq \frac{1}{2} \|\mathbf{v}\|_X^2 + \frac{V_{\alpha,\max}}{2\alpha} \|\mathbf{v}\|_X^2 + \frac{3\beta C_1^2}{2} \|\mathbf{u}\|_X^2 \|\mathbf{v}\|_X^2 + \beta C_1^2 \|\mathbf{u}\|_X \|\mathbf{v}\|_X^3 + \frac{\beta C_1^2}{4} \|\mathbf{v}\|_X^4.$$

Proof. It can be computed that

$$\mathbf{E}_h(\mathbf{u} + \mathbf{v}) - \mathbf{E}_h(\mathbf{u}) = \left(\frac{1}{2}\langle -\Delta_h(\mathbf{u}+\mathbf{v}), \mathbf{u}+\mathbf{v} \rangle_h - \frac{1}{2}\langle -\Delta_h\mathbf{u}, \mathbf{u} \rangle_h \right)$$
$$+ \left(\frac{1}{2}\langle \mathbb{V}(\mathbf{u}+\mathbf{v}), \mathbf{u}+\mathbf{v} \rangle_h - \frac{1}{2}\langle \mathbb{V}\mathbf{u}, \mathbf{u} \rangle_h \right)$$
$$+ \left(\frac{\beta}{4}\langle (\mathbf{u}+\mathbf{v})^2, (\mathbf{u}+\mathbf{v})^2 \rangle_h - \frac{\beta}{4}\langle \mathbf{u}^2, \mathbf{u}^2 \rangle_h \right)$$
$$= \langle -\Delta_h\mathbf{u} + \mathbb{V}\mathbf{u} + \beta\mathbf{u}^3, \mathbf{v} \rangle_h$$
$$+ \left(\frac{1}{2}\langle -\Delta_h\mathbf{v}, \mathbf{v} \rangle_h + \frac{1}{2}\langle \mathbb{V}\mathbf{v}, \mathbf{v} \rangle_h + \frac{3\beta}{2}\langle \mathbf{u}^2, \mathbf{v}^2 \rangle_h + \beta\langle \mathbf{u}, \mathbf{v}^3 \rangle_h + \frac{\beta}{4}\langle \mathbf{v}^2, \mathbf{v}^2 \rangle_h \right),$$

which leads to

$$|\mathbf{E}_h(\mathbf{u} + \mathbf{v}) - \mathbf{E}_h(\mathbf{u}) - \langle \nabla_X \mathbf{E}_h(\mathbf{u}), \mathbf{v} \rangle_X|$$
$$= \left| \mathbf{E}_h(\mathbf{u}+\mathbf{v}) - \mathbf{E}_h(\mathbf{u}) - \langle -\Delta_h\mathbf{u} + \mathbb{V}\mathbf{u} + \beta\mathbf{u}^3, \mathbf{v} \rangle_h \right|$$
$$\leq \frac{1}{2}\langle (-\Delta_h + \alpha\mathbb{I})\mathbf{v}, \mathbf{v} \rangle_h + \frac{1}{2}\langle (\mathbb{V}-\alpha\mathbb{I})\mathbf{v}, \mathbf{v} \rangle_h + \frac{3\beta}{2}\langle \mathbf{u}^2, \mathbf{v}^2 \rangle_h + \beta|\langle \mathbf{u}, \mathbf{v}^3 \rangle_h| + \frac{\beta}{4}\langle \mathbf{v}^2, \mathbf{v}^2 \rangle_h$$
$$\leq \frac{1}{2}\|\mathbf{v}\|_X^2 + \frac{V_{\alpha,\max}}{2}\|\mathbf{v}\|_2^2 + \frac{3\beta}{2}\|\mathbf{u}^2\|_2\|\mathbf{v}^2\|_2 + \beta\|\mathbf{u}^2\|_2^{\frac{1}{2}}\|\mathbf{v}^2\|_2^{\frac{3}{2}} + \frac{\beta}{4}\|\mathbf{v}^2\|_2^2$$
$$\leq \frac{1}{2}\|\mathbf{v}\|_X^2 + \frac{V_{\alpha,\max}}{2\alpha}\|\mathbf{v}\|_X^2 + \frac{3\beta C_1^2}{2}\|\mathbf{u}\|_X^2\|\mathbf{v}\|_X^2 + \beta C_1^2\|\mathbf{u}\|_X\|\mathbf{v}\|_X^3 + \frac{\beta C_1^2}{4}\|\mathbf{v}\|_X^4.$$

$\square$

We can now present the proof of Theorem 5.2 and Corollary 5.3.

Proof of Theorem 5.2. Let C_1 and C_2 be constants in Lemma 5.5. Define $C_3 \leq 1$ as a constant depending on α and $V_{\min}$ that satisfies

(5.3) $$\langle -\Delta_h\mathbf{u}, \mathbf{u} \rangle_h + \langle \mathbb{V}\mathbf{u}, \mathbf{u} \rangle_h \geq C_3\|\mathbf{u}\|_X^2, \quad \forall \mathbf{u} \in \mathbb{R}^N.$$

Define

$$C_u = \left(\left(\frac{2}{C_3} + \frac{2V_{\max}}{\alpha C_3} \right) \|\mathbf{u}^0\|_X^2 + \frac{\beta C_1^2}{2C_3} \|\mathbf{u}^0\|_X^4 \right)^{1/2}, \quad C_g = C_u + \frac{V_{\alpha,\max}}{\alpha}C_u + \frac{\beta C_2}{\sqrt{\alpha}}C_u^3.$$

We prove the theorem by induction. It is clear that (i) holds for $n = 0$ since $C_3 \leq 1$ implies $C_u \geq \|\mathbf{u}^0\|_X$. Suppose that (i) holds for $0, 1, \ldots, n$ and that (ii) and (iii) hold for $0, 1, \ldots, n-1$. We aim to show that (ii) and (iii) hold for n and that (i) holds for $n+1$.

It follows directly from Lemma 5.6 and (i) that (ii) holds for n. We focus on (iii) then. The iterative scheme is

$$\mathbf{u}^{n+1} = R_h\left(\mathbf{u}^n - \tau_n\nabla_X^{\mathcal{R}}\mathbf{E}_h(\mathbf{u}^n) \right) = \mathbf{u}^n - \tau_n\nabla_X^{\mathcal{R}}\mathbf{E}_h(\mathbf{u}^n) + \mathbf{R}^n,$$
$$\mathbf{R}^n = R_h\left(\mathbf{u}^n - \tau_n\nabla_X^{\mathcal{R}}\mathbf{E}_h(\mathbf{u}^n) \right) - \left(\mathbf{u}^n - \tau_n\nabla_X^{\mathcal{R}}\mathbf{E}_h(\mathbf{u}^n) \right).$$

According to Lemma 5.4 and Lemma 5.5, it holds that

$$\|\mathbf{R}^n\|_X \leq \frac{\tau_n^2}{2} \left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_2^2 \left\|\mathbf{u}^n - \tau_n \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X$$

$$\leq \frac{\tau_n^2}{2\alpha}(C_u + C_g) \left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X^2 \leq \frac{(C_u + C_g)C_g^2}{2\alpha},$$

where we used $\tau_n \leq 1$. Denote $\mathbf{g}^n = \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)$ and $\tilde{\mathbf{u}}^n = \mathbf{u}^n - \tau_n \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)$. By Lemma 5.7, we have

$$|E_h(\tilde{\mathbf{u}}^n) - E_h(\tilde{\mathbf{u}}^n + \mathbf{R}^n)| \leq |\langle \nabla_X E_h(\tilde{\mathbf{u}}^n), \mathbf{R}^n \rangle_X| + \left(\frac{1}{2} + \frac{V_{\alpha,\max}}{2\alpha}\right) \|\mathbf{R}^n\|_X^2$$

$$+ \left(\frac{3\beta C_1^2}{2} \|\tilde{\mathbf{u}}^n\|_X^2 \|\mathbf{R}^n\|_X^2 + \beta C_1^2 \|\tilde{\mathbf{u}}^n\|_X \|\mathbf{R}^n\|_X^3 + \frac{\beta C_1^2}{4} \|\mathbf{R}^n\|_X^4\right)$$

$$\leq \|\mathbf{R}^n\|_X \left(\|\nabla_X E_h(\tilde{\mathbf{u}}^n)\|_X + \left(\frac{1}{2} + \frac{V_{\alpha,\max}}{2\alpha}\right) \|\mathbf{R}^n\|_X\right.$$

$$\left. + \frac{3\beta C_1^2}{2} \|\tilde{\mathbf{u}}^n\|_X \|\mathbf{R}^n\|_X^2 + \beta C_1^2 \|\tilde{\mathbf{u}}^n\|_X \|\mathbf{R}^n\|_X^2 + \frac{\beta C_1^2}{4} \|\mathbf{R}^n\|_X^3\right)$$

$$\leq \tau_n^2 C_R \left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X^2,$$

where C_R depends only on C_u, C_g, $V_{\alpha,\max}$, α, β, and C_1. Similarly, we have

$$\left|E_h\left(\mathbf{u}^n - \tau_n \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right) - E_h(\mathbf{u}^n) - \left\langle \nabla_X E_h(\mathbf{u}^n), -\tau_n \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\rangle_X\right|$$

$$\leq \frac{1}{2} \|\tau_n \mathbf{g}^n\|_X^2 + \frac{V_{\alpha,\max}}{2\alpha} \|\tau_n \mathbf{g}^n\|_X^2 + \frac{3\beta C_1^2}{2} \|\mathbf{u}^n\|_X^2 \|\tau_n \mathbf{g}^n\|_X^2$$

$$+ \beta C_1^2 \|\mathbf{u}^n\|_X \|\tau_n \mathbf{g}^n\|_X^3 + \frac{\beta C_1^2}{4} \|\tau_n \mathbf{g}^n\|_X^4$$

$$\leq \tau_n^2 \left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X^2 \left(\frac{1}{2} + \frac{V_{\alpha,\max}}{2\alpha} + \frac{3\beta C_1^2 C_u^2}{2} + \beta C_1^2 C_u C_g + \frac{\beta C_1^2 C_g^2}{4}\right).$$

By Lemma 5.6, if $\tau_{\max}\left(\frac{1}{2} + \frac{V_{\alpha,\max}}{2\alpha} + \frac{3\beta C_1^2 C_u^2}{2} + \beta C_1^2 C_u C_g + \frac{\beta C_1^2 C_g^2}{4} + C_R\right) \leq \frac{1}{2}$, then

$$\mathbf{E}_h(\mathbf{u}^n) - \mathbf{E}_h(\mathbf{u}^{n+1}) = \mathbf{E}_h(\mathbf{u}^n) - E_h(\tilde{\mathbf{u}}^n) + E_h(\tilde{\mathbf{u}}^n) - E_h(\tilde{\mathbf{u}}^n + \mathbf{R}^n)$$

$$\geq \tau_n \left\langle \nabla_X \mathbf{E}_h(\mathbf{u}^n), \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\rangle_X - |E_h(\tilde{\mathbf{u}}^n) - E_h(\tilde{\mathbf{u}}^n + \mathbf{R}^n)|$$

$$- \left|E_h\left(\mathbf{u}^n - \tau_n \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right) - \mathbf{E}_h(\mathbf{u}^n) - \left\langle \nabla_X \mathbf{E}_h(\mathbf{u}^n), -\tau_n \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\rangle_X\right|$$

$$\geq \tau_n \left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X^2 - \tau_n^2 C_R \left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X^2$$

$$- \tau_n^2 \left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X^2 \left(\frac{1}{2} + \frac{V_{\alpha,\max}}{2\alpha} + \frac{3\beta C_1^2 C_u^2}{2} + \beta C_1^2 C_u C_g + \frac{\beta C_1^2 C_g^2}{4}\right)$$

$$\geq \frac{\tau_{\min}}{2} \left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X^2,$$

which means that (iii) holds for n. Since

$$\mathbf{E}_h(\mathbf{u}^{n+1}) \geq \frac{1}{2}\langle -\Delta_h \mathbf{u}^{n+1}, \mathbf{u}^{n+1}\rangle_h + \frac{1}{2}\langle \mathbb{V}\mathbf{u}^{n+1}, \mathbf{u}^{n+1}\rangle_h \geq \frac{1}{2}C_3\|\mathbf{u}^{n+1}\|_X^2,$$

one can conclude that

$$\left\|\mathbf{u}^{n+1}\right\|_X^2 \leq \frac{2}{C_3}\mathbf{E}_h(\mathbf{u}^{n+1}) \leq \frac{2}{C_3}\mathbf{E}_h(\mathbf{u}^0) \leq \frac{2}{C_3}\left\|\mathbf{u}^0\right\|_X^2 + \frac{2V_{\max}}{C_3}\left\|\mathbf{u}^0\right\|_2^2 + \frac{\beta}{2C_3}\left\|(\mathbf{u}^0)^2\right\|_2^2$$
$$\leq \left(\frac{2}{C_3} + \frac{2V_{\max}}{\alpha C_3}\right)\left\|\mathbf{u}^0\right\|_X^2 + \frac{\beta C_1^2}{2C_3}\left\|\mathbf{u}^0\right\|_X^4,$$

which implies that $\left\|\mathbf{u}^{n+1}\right\|_X \leq C_u$, i.e., (i) holds for $n+1$. Furthermore, we also have (ii) hold for $n+1$ by applying (5.2) and $\left\|\mathbf{u}^{n+1}\right\|_X \leq C_u$. $\qquad\square$

Proof of Corollary 5.3. It follows from Theorem 5.2 that $\lim_{n\to\infty}\left\|\nabla_X^{\mathcal{R}}\mathbf{E}_h(\mathbf{u}^n)\right\|_X = 0$. Let $\mathbf{u}^*$ be any limit point of $\{\mathbf{u}^n\}_{n=0}^\infty$, i.e., $\mathbf{u}^{n_l} \to \mathbf{u}^*$ for some subsequence. Then $\mathbb{G}_X\mathbf{u}^{n_l} \to \mathbb{G}_X\mathbf{u}^*$ and $\nabla_X\mathbf{E}_h(\mathbf{u}^{n_l}) \to \nabla_X\mathbf{E}_h(\mathbf{u}^*)$. Hence $\nabla_X^{\mathcal{R}}\mathbf{E}_h(\mathbf{u}^{n_l}) \to \nabla_X^{\mathcal{R}}\mathbf{E}_h(\mathbf{u}^*)$, which leads to $\nabla_X^{\mathcal{R}}\mathbf{E}_h(\mathbf{u}^*) = 0$. $\qquad\square$

5.2. **Locally exponential convergence.** In this subsection, we prove the local convergence rate of the modified H^1 scheme (4.4). We need Assumption 5.8.

Assumption 5.8. Let $\mathbf{u}^* \in \mathcal{M}$ be the ground state of the nonlinear eigenvalue problem (2.7), i.e., the global minimizer to (2.12). We assume that the multiplicity of the smallest eigenvalue of $A_{\mathbf{u}^*} = -\Delta_h + \mathbb{V} + \beta\mathrm{diag}(\mathbf{u}^*)^2$ is one, i.e., $\lambda_h^0 < \lambda_h^1$ where λ_h^0 and λ_h^1 be the smallest and the second smallest eigenvalue of $A_{\mathbf{u}^*}$.

Remark 5.9. Perron-Frobenius Theorem ensures Assumption 5.8 if a monotone scheme is used. For example, the matrix $A_{\mathbf{u}^*}$ is monotone for the second-order finite difference scheme with any mesh size h. For Q^2 spectral element method with a priori assumption on infinity norm of $\mathbf{u}^*$, the matrix $A_{\mathbf{u}^*}$ is monotone for small enough mesh size. However, Perron-Frobenius Theorem does not provide a quantification of the eigengap $|\lambda_h^0 - \lambda_h^1|$. In Appendix E, we prove that $|\lambda_h^0 - \lambda_h^1|$ has a uniform positive lower bound as $h \to 0$ for the P^1 finite element method with quadrature on an unstructured shape regular simplicial mesh.

The main local convergence result is stated as follows.

Theorem 5.10. *Suppose that Assumption* 5.1 *and Assumption* 5.8 *hold. Assume the step size bounds* $\tau_{\min}$ *and* $\tau_{\max}$ *satisfy*

$$(5.4) \qquad \sup_{\tau_{\min} \leq \tau \leq \tau_{\max}}\left\{(1 + L_g^2\tau^2) - \frac{\tau}{C_3}\min\left\{\frac{\lambda_h^1 - \lambda_h^0}{\lambda_h^0}, 1\right\}\right\} \leq C_\tau < 1,$$

where L_g *depends on* $\Omega, V_{\min}, V_{\max}, \alpha, \beta, k$ *and* $\mathbf{u}^*$. *Then* $\{\mathbf{u}^n\}_{n=0}^\infty$ *converges exponentially to the ground state* $\mathbf{u}^*$ *in* $\|\cdot\|_X$ *when* $\|\mathbf{u}^0 - \mathbf{u}^*\|_X$ *is sufficiently small.*

For given $L_g, C_3, \frac{\lambda_h^1 - \lambda_h^0}{\lambda_h^0} > 0$, the condition (5.4) holds for some $C_\tau < 1$ as long as $\tau_{\min} > 0$ and $\tau_{\max}$ is sufficiently small, since the coefficient of the linear term is negative.

Lemma 5.11. *Suppose that Assumption* 5.8 *holds. Then*

$$(5.5) \quad \mathbf{E}_h(\mathbf{u}) - \mathbf{E}_h(\mathbf{u}^*) \geq \frac{\lambda_h^1 - \lambda_h^0}{2}\|\mathbf{u} - \mathbf{u}^*\|_2^2 - \frac{\lambda_h^1 - \lambda_h^0}{8}\|\mathbf{u} - \mathbf{u}^*\|_2^4, \quad \forall\mathbf{u} \in \mathbb{R}^N.$$

Lemma 5.12. *Suppose that Assumption* 5.8 *holds. Then there exists constants* $L_\gamma, L_g > 0$ *depending on* $\Omega, V, \alpha, \beta, k$ *such that*

$$(5.6) \ \left\|\nabla_X^{\mathcal{R}}\mathbf{E}_h(\mathbf{u})\right\|_X \leq L_g\|\mathbf{u} - \mathbf{u}^*\|_X, \quad \left|\frac{\langle\mathbf{u}, \nabla_X\mathbf{E}_h(\mathbf{u})\rangle_h}{\langle\mathbb{G}_X\mathbf{u}, \mathbf{u}\rangle_h} - \lambda_h^0\right| \leq L_\gamma\|\mathbf{u} - \mathbf{u}^*\|_X,$$

as long as $\mathbf{u} \in \mathcal{M}$ *and* $\|\mathbf{u} - \mathbf{u}^*\|_X$ *is sufficiently small.*

The proof of the two lemmas will be given in Appendix D.

Proof of Theorem 5.10. If $\|\mathbf{u}^n - \mathbf{u}^*\|_X$ is sufficiently small, then by Lemma 5.12,

$$\left\|(\mathbf{u}^n - \mathbf{u}^*) - \tau_n \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X^2$$
$$= \|\mathbf{u}^n - \mathbf{u}^*\|_X^2 - 2\tau_n \langle \mathbf{u}^n - \mathbf{u}^*, \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\rangle_X + \tau_n^2 \left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X^2$$
$$\leq (1 + L_g^2 \tau_n^2) \|\mathbf{u}^n - \mathbf{u}^*\|_X^2 - 2\tau_n \langle \mathbf{u}^n - \mathbf{u}^*, \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\rangle_X$$
$$= (1 + L_g^2 \tau_n^2) \|\mathbf{u}^n - \mathbf{u}^*\|_X^2 + 2\tau_n \langle \mathbf{u}^* - \mathbf{u}^n, \nabla_X \mathbf{E}_h(\mathbf{u}^n)\rangle_X - 2\tau_n \gamma^n \langle \mathbf{u}^* - \mathbf{u}^n, \mathbb{G}_X \mathbf{u}^n\rangle_X,$$

where we used (4.4b) with $\gamma^n = \frac{\langle \mathbf{u}^n, \nabla_X \mathbf{E}_h(\mathbf{u}^n)\rangle_h}{\langle \mathbb{G}_X \mathbf{u}^n, \mathbf{u}^n\rangle_h}$. Set $\mathbf{e}^n = \mathbf{u}^n - \mathbf{u}^*$, then

$$\mathbf{E}_h(\mathbf{u}^*) - \mathbf{E}_h(\mathbf{u}^n) = \frac{1}{2}\langle -\Delta_h(\mathbf{u}^n - \mathbf{e}^n), \mathbf{u}^n - \mathbf{e}^n\rangle_h + \frac{1}{2}\langle \mathbb{V}(\mathbf{u}^n - \mathbf{e}^n), \mathbf{u}^n - \mathbf{e}^n\rangle_h$$

$$+ \frac{\beta}{4}\langle (\mathbf{u}^n - \mathbf{e}^n)^2, (\mathbf{u}^n - \mathbf{e}^n)^2\rangle_h - \frac{1}{2}\langle -\Delta_h \mathbf{u}^n, \mathbf{u}^n\rangle_h$$

$$- \frac{1}{2}\langle \mathbb{V}\mathbf{u}^n, \mathbf{u}^n\rangle_h - \frac{\beta}{4}\langle (\mathbf{u}^n)^2, (\mathbf{u}^n)^2\rangle_h$$

$$= -\langle \nabla_X \mathbf{E}_h(\mathbf{u}^n), \mathbf{e}^n\rangle_X + \frac{1}{2}\langle -\Delta_h \mathbf{e}^n, \mathbf{e}^n\rangle_h + \frac{1}{2}\langle \mathbb{V}\mathbf{e}^n, \mathbf{e}^n\rangle_h$$

$$+ \frac{3\beta}{2}\langle (\mathbf{u}^n)^2, (\mathbf{e}^n)^2\rangle_h - \beta\langle \mathbf{u}^n, (\mathbf{e}^n)^3\rangle_h + \frac{\beta}{4}\langle (\mathbf{e}^n)^2, (\mathbf{e}^n)^2\rangle_h$$

$$= -\langle \nabla_X \mathbf{E}_h(\mathbf{u}^n), \mathbf{e}^n\rangle_X + \frac{1}{2}\langle A_{\mathbf{u}^*} \mathbf{e}^n, \mathbf{e}^n\rangle_h$$

$$- \frac{\beta}{2}\langle (\mathbf{u}^*)^2, (\mathbf{e}^n)^2\rangle_h + \frac{3\beta}{2}\langle (\mathbf{u}^n)^2, (\mathbf{e}^n)^2\rangle_h - \beta\langle \mathbf{u}^n, (\mathbf{e}^n)^3\rangle_h + \frac{\beta}{4}\langle (\mathbf{e}^n)^2, (\mathbf{e}^n)^2\rangle_h$$

$$\geq -\langle \nabla_X \mathbf{E}_h(\mathbf{u}^n), \mathbf{e}^n\rangle_X + \frac{1}{2}\langle A_{\mathbf{u}^*} \mathbf{e}^n, \mathbf{e}^n\rangle_h$$

$$+ \frac{\beta}{2}\langle (\mathbf{u}^n)^2 - (\mathbf{u}^*)^2, (\mathbf{e}^n)^2\rangle_h - \beta\langle \mathbf{u}^n, (\mathbf{e}^n)^3\rangle_h + \frac{\beta}{4}\langle (\mathbf{e}^n)^2, (\mathbf{e}^n)^2\rangle_h$$

$$= -\langle \nabla_X \mathbf{E}_h(\mathbf{u}^n), \mathbf{e}^n\rangle_X + \frac{1}{2}\langle A_{\mathbf{u}^*} \mathbf{e}^n, \mathbf{e}^n\rangle_h - \frac{\beta}{4}\|(\mathbf{e}^n)^2\|_2^2$$

$$\geq -\langle \nabla_X \mathbf{E}_h(\mathbf{u}^n), \mathbf{e}^n\rangle_X + \frac{1}{2}\langle A_{\mathbf{u}^*} \mathbf{e}^n, \mathbf{e}^n\rangle_h - \frac{\beta C_1^2}{4}\|\mathbf{e}^n\|_X^4,$$

where Lemma 5.5 was used in the last step. As a consequence, we obtain that

$$\langle \nabla_X \mathbf{E}_h(\mathbf{u}^n), \mathbf{u}^* - \mathbf{u}^n\rangle_X \leq \mathbf{E}_h(\mathbf{u}^*) - \mathbf{E}_h(\mathbf{u}^n) - \frac{1}{2}\langle A_{\mathbf{u}^*} \mathbf{e}^n, \mathbf{e}^n\rangle_h + \frac{\beta C_1^2}{4}\|\mathbf{e}^n\|_X^4$$

$$\leq -\frac{\lambda_h^1 - \lambda_h^0}{2}\|\mathbf{e}^n\|_2^2 + \frac{\lambda_h^1 - \lambda_h^0}{8}\|\mathbf{e}^n\|_2^4 - \frac{1}{2}\langle A_{\mathbf{u}^*} \mathbf{e}^n, \mathbf{e}^n\rangle_h + \frac{\beta C_1^2}{4}\|\mathbf{e}^n\|_X^4$$

$$\leq -\frac{\lambda_h^1 - \lambda_h^0}{2}\|\mathbf{e}^n\|_2^2 - \frac{1}{2}\langle A_{\mathbf{u}^*} \mathbf{e}^n, \mathbf{e}^n\rangle_h + \left(\frac{\lambda_h^1 - \lambda_h^0}{8\alpha^2} + \frac{\beta C_1^2}{4}\right)\|\mathbf{e}^n\|_X^4,$$

where we used (5.5) and Lemma 5.5. Note that $\mathbf{u}^n \in \mathcal{M} \Rightarrow \|\mathbf{u}^n\|_2^2 = 1$, thus

$$\langle \mathbf{u}^* - \mathbf{u}^n, \mathbb{G}_X \mathbf{u}^n\rangle_X = \langle \mathbf{u}^* - \mathbf{u}^n, \mathbf{u}^n\rangle_h = \frac{1}{2}\left(\|\mathbf{u}^*\|_2^2 - \|\mathbf{u}^n\|_2^2 - \|\mathbf{u}^n - \mathbf{u}^*\|_2^2\right) = -\frac{1}{2}\|\mathbf{e}^n\|_2^2.$$

Combining all estimations above we obtain that

$$\left\|(\mathbf{u}^n - \mathbf{u}^*) - \tau_n \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X^2$$

$$\leq (1 + L_g^2 \tau_n^2) \|\mathbf{u}^n - \mathbf{u}^*\|_X^2 + 2\tau_n \langle \mathbf{u}^* - \mathbf{u}^n, \nabla_X \mathbf{E}_h(\mathbf{u}^n) \rangle_X - 2\tau_n \gamma^n \langle \mathbf{u}^* - \mathbf{u}^n, \mathbb{G}_X \mathbf{u}^n \rangle_X$$

$$\leq (1 + L_g^2 \tau_n^2) \|\mathbf{e}^n\|_X^2 + \tau_n (\gamma^n - (\lambda_h^1 - \lambda_h^0)) \|\mathbf{e}^n\|_2^2 - \tau_n \langle A_{\mathbf{u}^*} \mathbf{e}^n, \mathbf{e}^n \rangle_h$$

$$+ \tau_n \left(\frac{\lambda_h^1 - \lambda_h^0}{4\alpha^2} + \frac{\beta C_1^2}{2} \right) \|\mathbf{e}^n\|_X^4$$

$$\leq (1 + L_g^2 \tau_n^2) \|\mathbf{e}^n\|_X^2 + \tau_n (\lambda_h^0 - (\lambda_h^1 - \lambda_h^0)) \|\mathbf{e}^n\|_2^2 - \tau_n \langle A_{\mathbf{u}^*} \mathbf{e}^n, \mathbf{e}^n \rangle_h$$

$$+ \tau_n \frac{L_\gamma}{\alpha} \|\mathbf{e}^n\|_X^3 + \tau_n \left(\frac{\lambda_h^1 - \lambda_h^0}{4\alpha^2} + \frac{\beta C_1^2}{2} \right) \|\mathbf{e}^n\|_X^4$$

$$\leq (1 + L_g^2 \tau_n^2) \|\mathbf{e}^n\|_X^2 + \tau_n \max \left\{ \lambda_h^0 - (\lambda_h^1 - \lambda_h^0), 0 \right\} \frac{1}{\lambda_h^0} \langle A_{\mathbf{u}^*} \mathbf{e}^n, \mathbf{e}^n \rangle_h - \tau_n \langle A_{\mathbf{u}^*} \mathbf{e}^n, \mathbf{e}^n \rangle_h$$

$$+ \tau_n \frac{L_\gamma}{\alpha} \|\mathbf{e}^n\|_X^3 + \tau_n \left(\frac{\lambda_h^1 - \lambda_h^0}{4\alpha^2} + \frac{\beta C_1^2}{2} \right) \|\mathbf{e}^n\|_X^4$$

$$\leq (1 + L_g^2 \tau_n^2) \|\mathbf{e}^n\|_X^2 - \tau_n \min \left\{ \frac{\lambda_h^1 - \lambda_h^0}{\lambda_h^0}, 1 \right\} \langle A_{\mathbf{u}^*} \mathbf{e}^n, \mathbf{e}^n \rangle_h + \tau_n \frac{L_\gamma}{\alpha} \|\mathbf{e}^n\|_X^3$$

$$+ \tau_n \left(\frac{\lambda_h^1 - \lambda_h^0}{4\alpha^2} + \frac{\beta C_1^2}{2} \right) \|\mathbf{e}^n\|_X^4$$

$$\leq (1 + L_g^2 \tau_n^2) \|\mathbf{e}^n\|_X^2 - \frac{\tau_n}{C_3} \min \left\{ \frac{\lambda_h^1 - \lambda_h^0}{\lambda_h^0}, 1 \right\} \|\mathbf{e}^n\|_X^2$$

$$+ \tau_n \frac{L_\gamma}{\alpha} \|\mathbf{e}^n\|_X^3 + \tau_n \left(\frac{\lambda_h^1 - \lambda_h^0}{4\alpha^2} + \frac{\beta C_1^2}{2} \right) \|\mathbf{e}^n\|_X^4,$$

where we have used (5.6) and (5.3). Since we assume that $\|\mathbf{u}^n - \mathbf{u}^*\|_X$ is sufficiently small, there are some constants $C_u^{\mathrm{loc}}, C_g^{\mathrm{loc}} > 0$ such that $\|\mathbf{u}^n\|_X \leq C_u^{\mathrm{loc}}$ and $\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\|_X \leq C_g^{\mathrm{loc}}$. Recall in the proof of Theorem 5.2, we define

$$\mathbf{R}^n = R_h \left(\mathbf{u}^n - \tau_n \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n) \right) - \left(\mathbf{u}^n - \tau_n \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n) \right),$$

and have the following bound

$$\|\mathbf{R}^n\|_X \leq \frac{\tau_n^2}{2\alpha} (C_u^{\mathrm{loc}} + C_g^{\mathrm{loc}}) \left\|\nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X^2 \leq \frac{\tau_n^2}{2\alpha} (C_u^{\mathrm{loc}} + C_g^{\mathrm{loc}}) L_g^2 \|\mathbf{e}^n\|_X^2 .$$

With (5.4), we have

$$\left\|\mathbf{u}^{n+1} - \mathbf{u}^*\right\|_X \leq \left\|(\mathbf{u}^n - \mathbf{u}^*) - \tau_n \nabla_X^{\mathcal{R}} \mathbf{E}_h(\mathbf{u}^n)\right\|_X + \|\mathbf{R}^n\|_X$$

$$\leq \left((1 + L_g^2 \tau_n^2) \|\mathbf{e}^n\|_X^2 - \frac{\tau_n}{C_3} \min \left\{ \frac{\lambda_h^1 - \lambda_h^0}{\lambda_h^0}, 1 \right\} \|\mathbf{e}^n\|_X^2 + \tau_n \frac{L_\gamma}{\alpha} \|\mathbf{e}^n\|_X^3 \right.$$

$$\left. + \tau_n \left(\frac{\lambda_h^1 - \lambda_h^0}{4\alpha^2} + \frac{\beta C_1^2}{2} \right) \|\mathbf{e}^n\|_X^4 \right)^{1/2} + \frac{\tau_n^2}{2\alpha} (C_u^{\mathrm{loc}} + C_g^{\mathrm{loc}}) L_g^2 \|\mathbf{e}^n\|_X^2$$

$$\leq \left(C_\tau + \tau_n \frac{L_\gamma}{\alpha} \|\mathbf{e}^n\|_X + \tau_n \left(\frac{\lambda_h^1 - \lambda_h^0}{4\alpha^2} + \frac{\beta C_1^2}{2} \right) \|\mathbf{e}^n\|_X^2 \right)^{1/2} \|\mathbf{e}^n\|_X$$

$$+ \frac{\tau_n^2}{2\alpha} (C_u^{\mathrm{loc}} + C_g^{\mathrm{loc}}) L_g^2 \|\mathbf{e}^n\|_X^2 .$$

Since $C_\tau < 1$, if $\|\mathbf{e}^n\|_X$ is sufficiently small, then we get

$$\left\|\mathbf{u}^{n+1} - \mathbf{u}^*\right\|_X \leq C_\delta \left\|\mathbf{e}^n\right\|_X = C_\delta \left\|\mathbf{u}^n - \mathbf{u}^*\right\|_X,$$

where $C_\delta \in (0,1)$ is a constant. $\qquad\square$

At the end of this section, let us remark that our results for both global and local convergence are for $\alpha > 0$ and they do not directly apply to $\alpha = 0$ although the algorithm remains valid when setting $\alpha = 0$. The constants in Theorem 5.2 and Theorem 5.10 depend on the choice of $\alpha > 0$.

6. Numerical tests

For numerical implementation, we will only consider the Q^k spectral element method, which can be easily implemented with significant acceleration on modern GPUs [45]. In this section, we implement the H^1 gradient flow (4.4) with the following spatial discretization for the Laplacian operator and energy function:

(1) The Q^k spectral element method for any $k \geq 1$, see [42, 45].
(2) For $k = 1$, it is exactly the same as the classical second-order finite difference.
(3) The fourth-order compact finite difference scheme, see [40, 43] for details. The definition of the discrete energy is the same as the one for the second-order finite difference scheme, i.e., the trapezoidal rule is used for approximating the integral. Only the discrete Laplacian is replaced by the compact finite difference.

6.1. **Accuracy test of discrete Laplacian schemes.** We consider an exact solution to the nonlinear eigenvalue problem (1.1) on $\Omega = [-1,1]^3$ with a potential $V(\boldsymbol{x}) = \beta(1 - |u^*(\boldsymbol{x})|^2)$ where the ground state is

$$(6.1) \qquad u^*(\boldsymbol{x}) = \beta \sin\left(\pi\frac{x+1}{2}\right) \sin\left(\pi\frac{y+1}{2}\right) \sin\left(\pi\frac{z+1}{2}\right)$$

and the eigenvalue and energy are $\lambda^* = d\frac{\pi^2}{4} + \beta$, $\quad E(u^*) = \frac{1}{2}\lambda^* - \frac{\beta}{4}\left(\frac{3}{4}\right)^d$. We test the accuracy of various discrete Laplacian and discrete energy schemes, shown in Table 1. The H^1 scheme (4.4) converges within 20 iterations with $\alpha = 0.2$ and step size 1.

6.2. **Comparison of various gradient flow algorithms in 2D.** We consider the 2D problem with $V(\boldsymbol{x}) = \sin\left(\frac{\pi}{4}x\right)^2 \sin\left(\frac{\pi}{4}y\right)^2$ on the region $\Omega = [-16, 16]^2$. The performance of different gradient flow algorithms with fixed step size 1 is shown in Figure 1 (a) and (b). See [33] and references therein for the definition of these schemes. We emphasize that these algorithms could be faster with different step sizes, e.g., the L^2 flow will be faster with a larger step size, and A_u algorithm can be faster with adaptive step size. Here we just use the same step size to compare them. Notice that only $(-\Delta_h + \alpha I)^{-1}$ needs to be applied twice in the modified H^1 gradient flow (4.4). In each iteration of L^2, A_u and A_0 schemes, one needs to invert matrices like $-\Delta_h + V(\boldsymbol{x})$ or $-\Delta_h + V(\boldsymbol{x}) + \beta|\mathbf{u}|^2$, which is much more expensive than computing $(-\Delta_h + \alpha I)^{-1}$ [45]. As shown in Figure 1 (a) and (b), (4.4) with a proper parameter $\alpha > 0$ can allow a much larger step size for convergence, compared to $\alpha = 0$. Since H^1, A_0 and A_u schemes can all be written as Riemannian gradient descent methods, in each iteration, one can also numerically compute the best step size by minimizing the energy function w.r.t. the step size. For simplicity, we use

TABLE 1. Accuracy test of various schemes for a 3D problem (6.1). For both finite difference schemes, the discrete ground state coincides with the exact ground state at grid points for (6.1), so ground state errors are meaningless for (6.1) thus not listed.

The second-order finite difference (FD)					
FD grid	Δx	$\|\lambda_h^* - \lambda^*\|$	order	$\|E_h^* - E^*\|$	order
39^3	$h = 0.05$	3.80E-3	1.999	1.90E-3	-
79^3	$h = 0.025$	9.51E-4	2.000	4.76E-4	2.000
The fourth-order compact finite difference					
39^3	$h = 0.05$	1.17E-6	4.001	5.87E-7	-
79^3	$h = 0.025$	7.33E-8	4.000	3.66E-8	4.000

High-order finite element methods							
DoFs	Mesh	$\|\lambda_h^* - \lambda^*\|$	order	$\|E_h^* - E^*\|$	order	$\|\mathbf{u}_h^* - \mathbf{u}^*\|_\infty$	order
Q^2 spectral element method							
9^3	5^3	8.13E-4	-	4.05E-4	-	3.67E-4	-
19^3	10^3	5.02E-5	4.016	2.51E-5	4.012	2.39E-5	3.94
Q^3 spectral element method							
11^3	4^3	5.20E-5	-	2.87E-6	-	1.48E-5	-
23^3	8^3	8.98E-8	5.855	4.49E-8	6.000	5.21E-7	4.83
Q^4 spectral element method							
11^3	3^3	9.00E-7	-	5.22E-8	-	4.99E-6	-
23^3	6^3	4.11E-10	11.09	2.05E-10	7.990	8.30E-8	5.91
47^3	12^3	2.26E-12	7.506	1.23E-12	7.384	1.44E-9	5.85
Q^5 spectral element method							
9^3	2^3	1.49E-6	-	3.87E-5	-	1.14E-6	-
19^3	4^3	6.84E-11	14.41	6.50E-12	22.51	7.87E-9	7.17
Q^6 spectral element method							
9^3	2^3	8.08E-9	-	7.31E-7	-	4.06E-8	-
Q^7 spectral element method							
13^3	2^3	1.89E-10	-	9.99E-9	-	1.61E-9	-
Q^8 spectral element method							
15^3	2^3	3.76E-13	-	1.03E-10	-	5.76E-11	-

the *fminbnd* function in MATLAB to solve such a one-dimensional minimization problem, which involves evaluating the energy function quite a few times. In Figure 1 (c) and (d), we show the performance of H^1, A_0, and A_u schemes using such an optimal step size. We observe that iteration numbers in all the Riemannian gradient descent schemes with the optimal step size to reach convergence are almost the same. On the other hand, for CPU time of solving this particular problem, H^1 scheme with $\alpha = 0.15$ and the optimal step size is obviously slower than H^1 scheme with $\alpha = 0.15$ and a constant step size 1, due to the extra computational cost of computing the optimal step size, which would be more expensive for 3D problems.

6.3. **Comparison with the Backward Forward Euler method.** The modified H^1 flow has the advantage of inverting only constant coefficient Laplacian, which can be easily accelerated on modern GPUs as shown recently in [45]. On the other hand, in the literature, there are similar schemes, e.g., the Backward Forward Euler
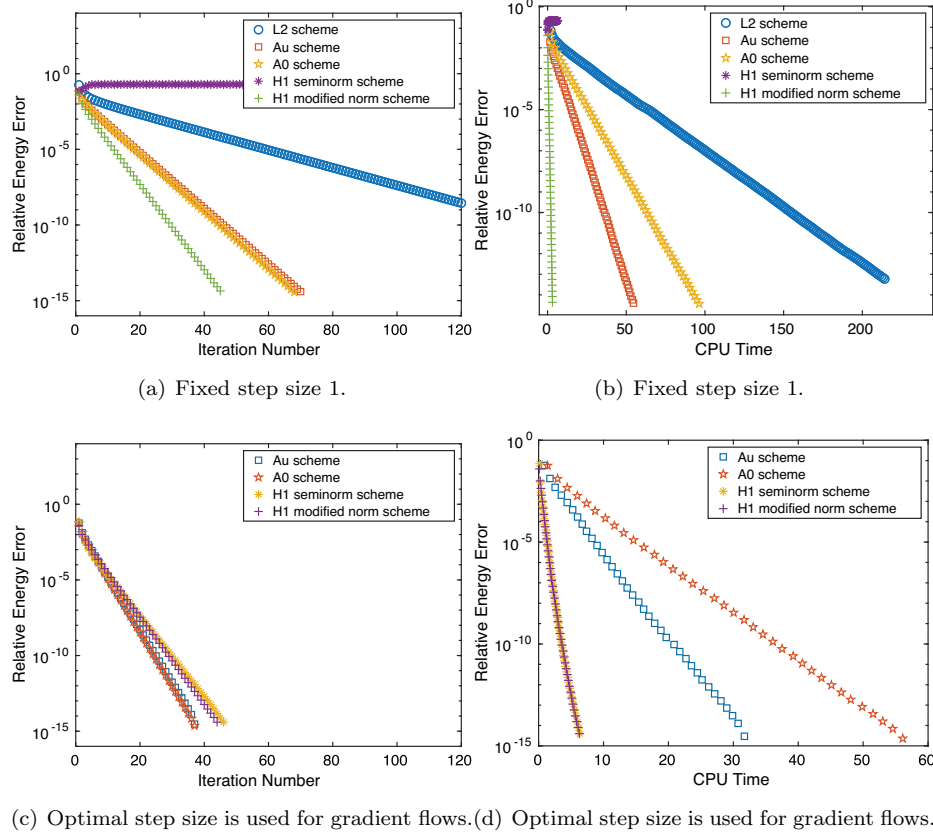
(a) Fixed step size 1.

(b) Fixed step size 1.

(c) Optimal step size is used for gradient flows. (d) Optimal step size is used for gradient flows.

FIGURE 1. A 2D example with $\beta = 10$ of second-order finite difference on a 800×800 grid. The modified H^1 norm has parameter $\alpha = 0.15$ in (4.4). The H^1 seminorm scheme is (4.4) with $\alpha = 0$. The CPU time for the H^1 scheme with $\alpha = 0.15$ and the fixed step size 1 to converge is 3 seconds, and the CPU time for the H^1 scheme with $\alpha = 0.15$ or $\alpha = 0$ with optimal step size to converge is more than 6 seconds. For L^2, A_u, and A_0 schemes (see [33] for definition), preconditioned conjugate gradient (PCG) is used for inverting a matrix like $-\Delta_h + V(\boldsymbol{x})$ and $(-\Delta_h)^{-1}$ is used as a preconditioner. The PCG converges within 30 iterations for all linear systems in this test.

method with a stabilization parameter in [7] is given by
(6.2)
$$\tilde{\mathbf{u}} = \left(-\Delta_h + \alpha \mathbb{I} + \frac{1}{\Delta t} \mathbb{I} \right)^{-1} \left(\alpha + \frac{1}{\Delta t} - \mathbb{V} - \beta \operatorname{diag}(\mathbf{u}^n)^2 \right) \mathbf{u}^n, \quad \mathbf{u}^{n+1} = \frac{\tilde{\mathbf{u}}}{\|\tilde{\mathbf{u}}\|_2}.$$

The modified H^1 flow is very different from (6.2). For instance, (4.4) is a Riemannian gradient descent method. In particular, only one inversion of Laplacian is needed in each step of (6.2), but there are two inversions of Laplacian in (4.4). The optimal parameter α for (6.2) was given in [7], yet it is unclear what the optimal

α should be for the modified H^1 flow. In numerical tests, the modified H^1 flow performs better when using tuned α, see Figure 2.
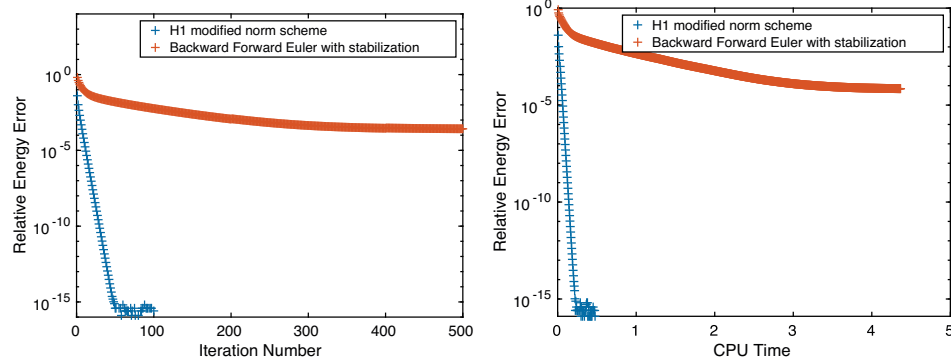

FIGURE 2. A 2D example with $\beta = 5$ of second-order discrete Laplacian on a 300×300 grid. The modified H^1 norm has parameter $\alpha = 0.15$ and step size 1 in (4.3). The Backward Forward Euler with stabilization (6.2) uses the optimal parameters α in [7] and the largest stable step size 0.1, which is also the most efficient step size. The initial condition is the ground state for $\beta = 0$.

6.4. **Tuning parameters.** We consider $\Omega = [-16, 16]^3$ with a potential:

$$(6.3) \qquad V(\boldsymbol{x}) = \sin\left(\frac{\pi}{4}x\right)^2 \sin\left(\frac{\pi}{4}y\right)^2 \sin\left(\frac{\pi}{4}z\right)^2 .$$

For (6.3) and $\beta = 10$, the modified H^1 scheme (4.4) with $\alpha = 0.15$ and $\tau = 1$ has the same convergence performance for any grid size or any discrete Laplacian, unless it is an extremely coarse grid, as shown in Figure 3 (a). Thus we can easily find the best step size for a given $\beta = 4000$ by tuning it on a 100^3 grid as shown in Figure 3 (b).

6.5. **3D implementation on GPUs.** As shown recently in [45], any discrete Laplacian on a Cartesian grid can be easily accelerated on modern GPUs with a simple implementation in MATLAB 2023. In particular, to invert a discrete Laplacian on a grid size 1000^3, it only takes 0.8 seconds on one Nividia A100 GPU card. And such a result holds for Q^k spectral element method, see [45] for details.

We consider solving the 3D problem with potential (6.3). See Figure 4 for visualization of the potential and its ground state for $\beta = 10$ and $\beta = 4000$. We define the online computation time as the computational time without counting the offline computational time such as preparing discrete Laplacian matrices and loading data to the GPU. For the potential 4 with $\beta = 10$, the modified H^1 flow scheme (4.4) with $\alpha = 0.15$ and $\tau = 1$, we stop the iteration when the relative residue stops decreasing, where the relative residue is defined as

$$\text{residue} = \left\| \mathbf{u} - \frac{-\Delta_h \mathbf{u} + \mathbb{V}\mathbf{u} + \beta|\mathbf{u}|^3}{\| -\Delta_h \mathbf{u} + \mathbb{V}\mathbf{u} + \beta|\mathbf{u}|^3 \|} \right\| / \|\mathbf{u}\|.$$

The results of computing energy and the eigenvalue are listed in Table 2, in which GPU time is the online computational time. In particular, the online computation
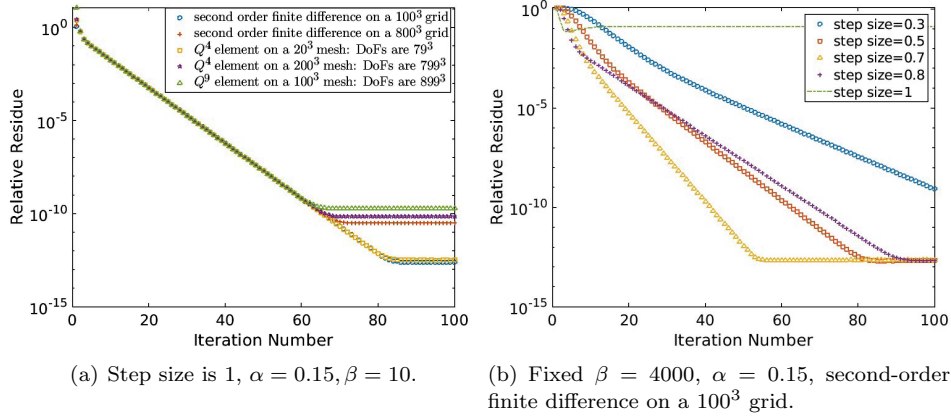
(a) Step size is 1, $\alpha = 0.15, \beta = 10$.

(b) Fixed $\beta = 4000$, $\alpha = 0.15$, second-order finite difference on a 100^3 grid.

FIGURE 3. The performance of the modified H^1 scheme (4.4) solving 3D Gross-Pitaevskii nonlinear eigenvalue problem with potential (6.3). The left shows that the performance is independent of discretization and grid size. Thus parameters can be tuned on a coarse grid as shown in the right.

time is 214 seconds on the Nvidia A100 for 100 iterations of (4.4) on a 1000^3 grid. For $\beta = 4000$, we use step size 0.7 and $\alpha = 0.15$, we stop the iteration when the relative residue stops decreasing. The performance is listed in Table 3. In both Tables 2 and 3, the reference solutions are generated by Q^{10} spectral element method on a 100^3 mesh, and the ground state errors $\frac{\|\mathbf{u}^* - \mathbf{u}_{ref}\|_{\ell\infty}}{\|\mathbf{u}_{ref}\|_\infty}$ are measured only at the nodes of matching with nodes of Q^{10} elements a 100^3 mesh. For instance, for Q^5 element on a 20^3 mesh, $\frac{\|\mathbf{u}^* - \mathbf{u}_{ref}\|_{\ell\infty}}{\|\mathbf{u}_{ref}\|_\infty}$ is measured only at the cell vertices.

(a) The potential function.
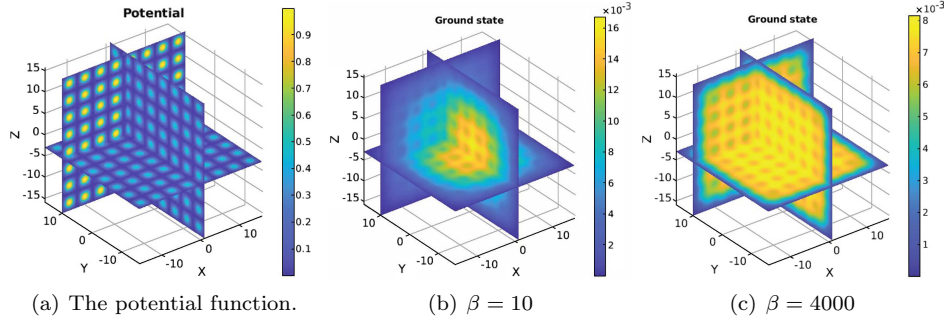
(b) $\beta = 10$

(c) $\beta = 4000$

FIGURE 4. The potential function (6.3) and its ground state

6.6. A 3D example with a combined harmonic and optical lattice potential.
We consider the 3D example in [7] with the following combined harmonic and optical lattice potential on the domain $[-8, 8]^3$:

$$V(x, y, z) = x^2 + y^2 + z^2 + 100 \left(\sin^2 \frac{\pi x}{4} + \sin^2 \frac{\pi y}{4} + \sin^2 \frac{\pi z}{4} \right).$$

TABLE 2. The 3D problem as shown in Figure 4 with $\beta = 10$. The GPU (Nvidia A100 80G) online computation time of the scheme (4.4) with $\alpha = 0.15$ and $\tau = 1$. The iteration stops when the relative residue stops decreasing. The reference eigenvalue $\lambda_{ref}^* = 0.143834048046$, energy $E_{ref} = 0.071660785256$ and ground state $\mathbf{u}_{ref}$ are generated by Q^{10} SEM on a 100^3 mesh.

The second-order finite difference					
DoFs	Mesh size	$\frac{\|\lambda_h^* - \lambda_{ref}^*\|}{\lambda_{ref}^*}$	$\frac{\|E_h(\mathbf{u}^*) - E_{ref}\|}{E_{ref}}$	$\frac{\|\mathbf{u}^* - \mathbf{u}_{ref}\|_{\ell\infty}}{\|\mathbf{u}_{ref}\|_\infty}$	Iteration #	GPU time
199^3	$h = \frac{32}{200}$	3.70E-4	3.72E-4	8.04E-4	86	0.88 second
999^3	$h = \frac{32}{1000}$	1.48E-5	1.48E-5	3.21E-5	76	165 seconds
The fourth-order compact finite difference						
199^3	$h = \frac{32}{200}$	1.17E-6	1.18E-6	2.00E-6	84	0.92 second
999^3	$h = \frac{32}{1000}$	1.86E-9	1.87E-9	3.21E-9	73	161 seconds
Q^4 spectral element method						
DoFs	FEM Mesh	$\frac{\|\lambda_h^* - \lambda_{ref}^*\|}{\lambda_{ref}^*}$	$\frac{\|E_h(\mathbf{u}^*) - E_{ref}\|}{E_{ref}}$	$\frac{\|\mathbf{u}^* - \mathbf{u}_{ref}\|_{\ell\infty}}{\|\mathbf{u}_{ref}\|_\infty}$	Iteration #	GPU time
199^3	50^3	8.57E-10	8.62E-10	2.73E-7	83	0.88 second
399^3	100^3	3.60E-12	3.62E-12	4.86E-9	78	6.10 seconds
Q^5 spectral element method						
99^3	20^3	3.53E-9	3.49E-9	6.06E-7	88	0.54 second
Q^{20} spectral element method						
99^3	5^3	5.31E-12	5.23E-12	1.29E-8	83	0.54 second
199^3	10^3	8.79E-12	8.85E-12	6.42E-12	75	0.82 second

TABLE 3. The 3D problem as shown in Figure 4 with $\beta = 4000$. The GPU (Nvidia A100 80G) online computation time of the scheme (4.4) with $\alpha = 0.15$ and $\tau = 0.7$. The iteration stops when the relative residue stops decreasing. The reference eigenvalue $\lambda_{ref}^* = 0.34919956116$, energy $E_{ref} = 0.127936543199$ and ground state $\mathbf{u}_{ref}$ are generated by Q^{10} SEM on a 100^3 mesh.

The second-order finite difference					
DoFs	Mesh size	$\frac{\|\lambda_h^* - \lambda_{ref}^*\|}{\lambda_{ref}^*}$	$\frac{\|E_h(\mathbf{u}^*) - E_{ref}\|}{E_{ref}}$	$\frac{\|\mathbf{u}^* - \mathbf{u}_{ref}\|_{\ell\infty}}{\|\mathbf{u}_{ref}\|_\infty}$	Iteration #	GPU time
999^3	$h = \frac{32}{1000}$	5.67E-6	7.82E-6	2.96E-5	48	105 seconds
The fourth-order compact finite difference						
999^3	$h = \frac{32}{1000}$	5.67E-10	8.72E-10	2.77E-9	48	106 seconds
Q^4 spectral element method						
DoFs	FEM Mesh	$\frac{\|\lambda_h^* - \lambda_{ref}^*\|}{\lambda_{ref}^*}$	$\frac{\|E_h(\mathbf{u}^*) - E_{ref}\|}{E_{ref}}$	$\frac{\|\mathbf{u}^* - \mathbf{u}_{ref}\|_{\ell\infty}}{\|\mathbf{u}_{ref}\|_\infty}$	Iteration #	GPU time
199^3	50^3	2.64E-10	4.06E-10	2.81E-7	54	0.63 second
399^3	100^3	1.79E-12	2.33E-12	4.45E-9	50	3.88 seconds
Q^5 spectral element method						
99^3	20^3	7.49E-8	6.44E-8	2.18E-7	57	0.42 second
Q^{20} spectral element method						
99^3	5^3	3.43E-12	4.07E-12	8.32E-9	54	0.37 second
199^3	10^3	3.58E-12	4.96E-12	3.22E-12	50	0.57 second

For $\beta = 1600$, we find that $\alpha = 10$ and $\Delta t = 0.1$ are efficient parameters. The modified H^1 flow with Q^{40} spectral element method on a $5 \times 5 \times 5$ finite element mesh converges with residue reaching 6.3×10^{-13} after 665 iterations using a simple and crude initial guess $\mathbf{u}^0 \equiv 1$. The online computation time is 6 seconds on Nvidia A100. The numerical ground state energy and eigenvalue are $E(\mathbf{u}_h^*) = 33.80227900547$ and $\lambda_h^* = 80.89511440602$. The results are consistent with findings in [7]. Due to the different definitions of energy functions in this paper and [7], $\beta = 1600$, $E(\mathbf{u}_h^*) = 33.80227900547$ and $\frac{1}{2}\lambda_h^* = 40.44755720301$ in this paper, correspond to the case for $\beta = 800$, $E_g = 33.8023$, and $\mu_g = 40.4476$ in [7]. See Figure 5.
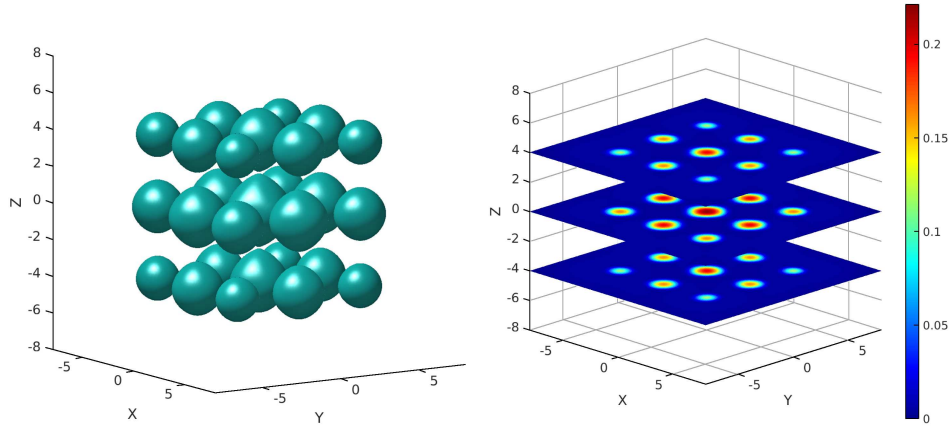

FIGURE 5. A 3D example for a combined harmonic and optical lattice potential. Left is the isosurface of the ground state for isovalue 0.002, and right is the slice view of the ground state. For $\beta = 1600$, using Q^{40} spectral element method on a 5^3 mesh, (4.4) with $\alpha = 10$ and $\tau = 0.1$ and $\mathbf{u}^0 \equiv 1$ converges after 665 iterations. The online computation time is 6 seconds on Nvidia A100. $E(\mathbf{u}_h^*) = 33.80227900547$ and $\lambda_h^* = 80.89511440602$, consistent with the results in [7].

7. Concluding remarks

We have considered the H^1 Sobolev gradient flow for finding the ground state of the Gross-Pitaevskii eigenvalue problem, under a modified H^1-norm. Global convergence to a critical point and the local exponential convergence rate have been established. Numerical experiments suggest that the scheme with the spectral element method can be very efficient when using tuned parameters, which can be easily and efficiently implemented on modern GPUs.

Appendix A. Explicit finite difference formulation and discrete energy

We give explicit equivalent finite difference formulation of the Q^k spectral element scheme (2.7), especially the Q^1 case, which is equivalent to the second-order

finite difference. We also give the explicit expressions of the P^1 finite element method on an unstructured mesh.

A.1. **The second-order finite difference scheme.** For a one-dimensional uniform grid $-L = x_0 < x_1 < \cdots < x_n < x_{n+1} = L$ with grid spacing $h = \frac{2L}{n+1}$, for any vector $u_h \in V_0^h$ with P^1 polynomial basis, it can be represented by $\mathbf{u} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}^\top$ with $u_0 = u_{n+1} = 0$. The discrete inner product is given by

$$\langle u_h, v_h \rangle = \frac{1}{2} h u_0 v_0 + \sum_{i=1}^{n} u_i v_i h + \frac{1}{2} h u_n v_n = h \mathbf{u}^\top \mathbf{v}.$$

Define $M = h \mathbb{I}_{n \times n}$ and $K = M^{-1} S$ where $S = \frac{1}{h} \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}_{n \times n}$.

Then the matrices in (2.6) and (2.7) for 1D are $\mathbb{M} = M$, $\mathbb{S} = S$, and $-\Delta_h = K$.

For a two-dimensional problem on a uniform grid for the domain $[-L, L]^2$, assume there are $n \times n$ interior grid points. Let U, V and F denote 2D arrays of size $n \times n$ consisting of point values of $u_h(x_1, x_2), V(x_1, x_2), f(x_1, x_2)$ at grid points. Let $\text{vec}(U)$ be the vector generated by arranging U column by column. The scheme (2.6) becomes

$$[S \otimes M + M \otimes S + \text{diag}(\text{vec}(V))]\text{vec}(U) + \beta M \otimes M \text{vec}(U^3) = \lambda_h M \otimes M \text{vec}(U),$$

where U^3 denotes the entrywise cubic power. In 2D, (2.7) can be written as

$$[K \otimes I + I \otimes K + \text{diag}(\text{vec}(V))]\text{vec}(U) + \beta \text{vec}(U^3) = \lambda_h \text{vec}(U).$$

With the property $(B^\top \otimes A)\text{vec}(X) = \text{vec}(AXB)$, it can be equivalently expressed as

$$KU + UK^\top + V \circ U + \beta U^3 = \lambda_h U,$$

where $\circ$ represents Hadamard product, i.e., entrywise product. With similar notation as in [45], the three-dimensional case of (2.7) can be expressed as

$$[K \otimes I \otimes I + I \otimes K \otimes I + I \otimes I \otimes K + \text{diag}(\text{vec}(V))]\text{vec}(U) + \beta \text{vec}(U^3) = \lambda_h \text{vec}(U).$$

Thus, the matrix $-\Delta_h$ is given explicitly as follows:

$$(A.1) \qquad -\Delta_h = \begin{cases} K, & d = 1, \\ K \otimes I + I \otimes K, & d = 2, \\ K \otimes I \otimes I + I \otimes K \otimes I + I \otimes I \otimes K, & d = 3. \end{cases}$$

Remark A.1. The Q^1 scheme with quadrature gives exactly the same second-order centered difference for the interior grid. For Neumann boundary condition, the Q^1 scheme with quadrature gives a slightly different scheme from a conventional finite difference scheme, see Remark 3.3 in [34]. When deriving finite difference from the finite element method, convergence is trivially implied by finite element error estimates.

A.2. The P^1 finite element scheme on a simplicial mesh. Let $\Omega_h \subset \mathbb{R}^d$ be a simplicial mesh satisfying (3.2). Let $\boldsymbol{x}_i, i = 1, \cdots, N$ denote all interior quadrature points. For any $u_h, v_h \in V_0^h$, since ∇u_h and ∇v_h are piecewise constants, the quadrature is exact. For an edge E in a simplex T, let $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ be its ends and $\delta_E u_h$ denote $u_h(\boldsymbol{x}_i) - u_h(\boldsymbol{x}_j) = u_i - u_j$. By formulae in [52, Section 2],

$$\langle \nabla u_h, \nabla v_h \rangle = \int_{\Omega_h} \nabla u_h \cdot \nabla v_h \mathrm{d}\boldsymbol{x} = \sum_{T \in \Omega_h} \sum_{E \subset T} \frac{1}{d(d-1)} |\kappa_E^T| \cot \theta_E^T \delta_E u_h \delta_E v_h,$$

which can also be written as

$$(A.2) \quad \langle \nabla u_h, \nabla v_h \rangle = \sum_E \sum_{T \supset E} \frac{1}{d(d-1)} |\kappa_E^T| \cot \theta_E^T (u_i - u_j)(v_i - v_j),$$

$$\boldsymbol{x}_i, \boldsymbol{x}_j \text{ are two ends of the edge } E.$$

By notation in Section 2, we have $\mathbf{u}^T \mathbb{S} \mathbf{v} = \langle \nabla u_h, \nabla v_h \rangle$, which implies that the stiffness matrix can be explicitly written as follows. The off-diagonal entries of $\mathbb{S}$ are given as

$$\mathbb{S}_{ij} = -\sum_{T \supset E} \frac{1}{d(d-1)} |\kappa_E^T| \cot \theta_E^T, \quad i \neq j \quad \text{and } E \text{ is the edge connecting } \boldsymbol{x}_i \text{ and } \boldsymbol{x}_j.$$

And the diagonal entries of $\mathbb{S}$ can be obtained by the fact that each row sum of $\mathbb{S}$ should be zero.

In a triangular mesh in two dimensions, for the edge connecting two interior vertices $\boldsymbol{x}_i, \boldsymbol{x}_j$, there are two angles θ_{ij}^1 and θ_{ij}^2 as shown in the figure, thus the stiffness matrix can be written as

$$\mathbb{S}_{ij} = \begin{cases} -\frac{\cot \theta_{ij}^1 + \cot \theta_{ij}^2}{2}, & j \neq i, \\ -\sum_{j \neq i} \mathbb{S}_{ij} & j = i. \end{cases}$$
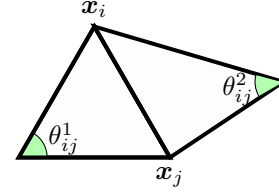
The necessary and sufficient condition for $\cot \theta_{ij}^1 + \cot \theta_{ij}^2 \geq 0$ is $\theta_{ij}^1 + \theta_{ij}^2 \leq \pi$. For $\mathbb{S}$ to be an M-matrix, it suffices to have $\theta_{ij}^1 + \theta_{ij}^2 \leq \pi$, which can be achieved in a Delaunay triangulation.

A.3. The discrete Laplacian from Q^k scheme. The full details can be found in [45]. Let S be the stiffness matrix and M be the mass matrix Q^k spectral element method in one dimension. In two dimensions (2.6) can be written as

$$[S \otimes M + M \otimes S + (M \otimes M) \operatorname{diag}(\operatorname{vec}(V))] \operatorname{vec}(U) = \lambda_h (M \otimes M) \operatorname{vec}(U).$$

Define $H = M^{-1}S$, the scheme (2.7) in two dimensions can be written as

$$[H \otimes I + I \otimes H + \operatorname{diag}(\operatorname{vec}(V))] \operatorname{vec}(U) + \beta \operatorname{vec}(U^3) = \lambda_h \operatorname{vec}(U),$$

or equivalently $HU + UH^\top + V \circ U + \beta U^3 = \lambda_h U$, and in three dimensions it is

$$[H \otimes I \otimes I + I \otimes H \otimes I + I \otimes I \otimes H + \operatorname{diag}(\operatorname{vec}(V))] \operatorname{vec}(U) + \beta \operatorname{vec}(U^3) = \lambda_h \operatorname{vec}(U).$$

It is possible to derive explicit entries of matrices S, M, H, see [42, 49] for more details.

Remark A.2. For $k \geq 2$, the discrete Laplacian above gives a $(k+2)$-th order finite difference scheme in discrete l^2-norm for solving elliptic equations [42] and for parabolic, wave and Schrödinger equations [39]. For solving a linear eigenvalue problem, e.g., $\beta = 0$ in (2.7), standard *a priori* error estimate for eigenvalues is that Q^k spectral element method gives $2k$-th order of accuracy if assuming sufficient regularity.

A.4. **Discrete energy of the second-order finite difference.** The formula (A.2) can be written out more explicitly when the vertices in the simplicial mesh form a Cartesian grid, i.e., when the P^1 finite element method becomes the second-order finite difference method. For the one-dimensional case, recall that $u_0 = u_{n+1} = 0$ for $u_h \in V_0^h$, we have

$$\forall u_h, v_h \in V_0^h, \quad \langle \nabla u_h, \nabla v_h \rangle = h \sum_{i=1}^{n+1} \frac{u_i - u_{i-1}}{h} \cdot \frac{v_i - v_{i-1}}{h}.$$

Define the matrix $D = \frac{1}{h} \begin{pmatrix} 1 & & & & \\ -1 & 1 & & & \\ & \ddots & \ddots & & \\ & & -1 & 1 \\ & & & -1 \end{pmatrix}_{(n+1) \times n}$. Then it satisfies $D^\top D =$

K. In one dimension $\langle u_h, u_h \rangle = \mathbf{u}^\top \mathbb{M} \mathbf{v} = h \mathbf{u}^\top \mathbf{v}$. Thus we have

$$\langle \nabla u_h, \nabla v_h \rangle = \langle D\mathbf{u}, D\mathbf{v} \rangle_h = \langle D^\top D\mathbf{u}, \mathbf{v} \rangle_h = \langle K\mathbf{u}, \mathbf{v} \rangle_h = \langle -\Delta_h \mathbf{u}, \mathbf{v} \rangle_h.$$

In two dimensions, by plugging in the quadrature, for any $u_h, v_h \in V_0^h$, we have
(A.3)

$$\langle \nabla u_h, \nabla v_h \rangle = h^2 \sum_{i,j=1}^{n+1} \frac{u_{i,j} - u_{i-1,j}}{h} \cdot \frac{v_{i,j} - v_{i-1,j}}{h} + \frac{u_{i,j} - u_{i,j-1}}{h} \cdot \frac{v_{i,j} - v_{i,j-1}}{h}.$$

With our notation for the two-dimensional problem, let $\mathbf{u} = \text{vec}(U)$, then we have

$$\langle \nabla u_h, \nabla v_h \rangle = \langle DU, DV \rangle_h + \langle UD, VD \rangle_h = \langle D^\top DU, V \rangle_h + \langle UDD^\top, V \rangle_h,$$

$$\langle \nabla u_h, \nabla v_h \rangle = \langle KU + UK, V \rangle_h = \langle (K \otimes I + I \otimes K)\text{vec}(U), \text{vec}(V) \rangle_h = \langle -\Delta_h \mathbf{u}, \mathbf{v} \rangle_h.$$

The three-dimensional case of the discrete gradient can be similarly written out.

APPENDIX B. M-MATRIX AND PERRON-FROBENIUS THEOREM

B.1. **M-matrix.** Nonsingular M-matrices are monotone. There are many equivalent definitions or characterizations of M-matrices, see [47]. The following is a convenient sufficient but not necessary characterization of nonsingular M-matrices [41]:

Theorem B.1. *For a real square matrix A with positive diagonal entries and non-positive off-diagonal entries, A is a nonsingular M-matrix if all the row sums of A are nonnegative and at least one row sum is positive.*

B.2. **Irreducible nonnegative matrices.** A matrix $A \in \mathbb{C}^{n \times n}$ is called *reducible* if there exists a permutation matrix P such that $PAP^\top$ is block upper triangular. A matrix is irreducible if and only the graph it represents is strongly connected.

Lemma B.2. *For a nonsingluar irreducible matrix A, A^{-1} is also irreducible.*

The following results can be found in [51]:

Theorem B.3 (Perron-Frobenius). *If $A \geq 0$ is irreducible, then:*

(1) *The spectral radius $\rho(A)$ is a simple eigenvalue of A with an eigenvector $x > 0$.*

(2) *$\rho(A)$ increases when any entry of A increases.*

Theorem B.4. *The positive eigenvector (Perron-Frobenius eigenvector) for an irreducible nonnegative matrix is unique up to scalar multiplication.*

Proof. Let $x > 0$ be the left Perron-Frobenius eigenvector then $x^\top A = \rho(A)x^\top$. If there exists another eigenvector $y > 0$ for an eigenvalue λ, then $Ay = \lambda y \implies x^\top Ay = \lambda x^\top y$. Since $x^\top A = \rho(A)x^\top \implies x^\top Ay = \rho(A)x^\top y$, we get $(\rho(A) - \lambda)x^\top y = 0$ and $x^\top y > 0 \implies \rho(A) = \lambda$. Thus there is only one eigenvalue, i.e., $\rho(A)$, with positive eigenvectors, and $\rho(A)$ is a simple eigenvalue by Theorem B.3. $\square$

APPENDIX C. DEFERRED PROOFS FOR SECTION 5.1

C.1. **Norm equivalence and standard regularity results.**

Lemma C.1. *There are positive constants D_1, D_2, D_3 independent of mesh size h, such that the following holds for any $v_h \in V_0^h$:*

$$(C.1) \qquad \frac{1}{D_p}|\langle v_h^p, v_h^p\rangle|^{\frac{1}{2p}} \leq \|v_h\|_{L^{2p}(\Omega)} \leq D_p|\langle v_h^p, v_h^p\rangle|^{\frac{1}{2p}}, \quad p = 1, 2, 3.$$

Proof. We first prove it for Q^k spectral element method. Consider a cubic cell $e = [x_1^e - \frac{h}{2}, x_1^e + \frac{h}{2}] \times \cdots \times [x_d^e - \frac{h}{2}, x_d^e + \frac{h}{2}] \in \Omega_h$ and a reference cell $\hat{K} = [-1,1]^3$. For $v_h(\mathbf{x})$ defined on e, consider $\hat{v}_h(\mathbf{t}) = v_h\left(t_1\frac{h}{2} + x_1^e, \cdots, t_d\frac{h}{2} + x_d^e\right)$, which is defined on $\hat{K}$. Let $\langle \hat{v}_h, \hat{v}_h\rangle_{\hat{K}}$ denote the approximation to the integral $\int_{\hat{K}} |\hat{v}_h(\mathbf{t})|^2 d\mathbf{t}$ by $(k+1)$-point Gauss Lobatto quadrature for each variable. Since both $\sqrt{\langle \hat{v}_h, \hat{v}_h\rangle_{\hat{K}}}$ and $\sqrt{\int_{\hat{K}} |\hat{v}_h(\mathbf{t})|^2 d\mathbf{t}}$ are norms of $Q^k(\hat{K})$, by the equivalence of any two norms on the finite-dimensional space $Q^k(\hat{K})$, we have

$$\forall \hat{v}_h \in Q^k(\hat{K}), \quad \frac{1}{D_1}\sqrt{\langle \hat{v}_h, \hat{v}_h\rangle_{\hat{K}}} \leq \sqrt{\int_{\hat{K}} |\hat{v}_h(\mathbf{t})|^2 d\mathbf{t}} \leq D_1\sqrt{\langle \hat{v}_h, \hat{v}_h\rangle_{\hat{K}}}.$$

By mapping back to e, and summing over e, we get (C.1) for $p = 1$.

With the same notation and arguments above, for a Q^k polynomial v_h with $(k+1)$-point Gauss-Lobatto quadrature, let $\hat{w}_i$ and $\hat{v}_i$ be quadrature weights and quadrature node values on the reference cell $\hat{K}$, then we have $|\langle \hat{v}_h^2, \hat{v}_h^2\rangle_{\hat{K}}|^{\frac{1}{4}} = \left|\sum_{i=1}^{(k+1)^d} \hat{w}_i \hat{v}_i^4\right|^{\frac{1}{4}}$, which can be easily verified to be a norm of $\mathbb{R}^{(k+1)^d}$ thus a norm of $Q^k(\hat{K})$. By the equivalence of any two norms on $Q^k(\hat{K})$, we have

$$\forall \hat{v}_h \in Q^k(\hat{K}), \quad \frac{1}{D_2}\left|\langle \hat{v}_h^2, \hat{v}_h^2\rangle_{\hat{K}}\right|^{\frac{1}{4}} \leq \|\hat{v}_h\|_{L^4(\hat{K})} \leq D_2\left|\langle \hat{v}_h^2, \hat{v}_h^2\rangle_{\hat{K}}\right|^{\frac{1}{4}}.$$

By mapping back to e, and summing over e, we get (C.1) for $p = 2$. The proof of $p = 3$ is almost identical to the case $p = 3$. Finally, the same proof also applies to P^k finite element method on a simplicial mesh with quadrature. $\square$

Second, we want to show the X-norm is equivalent to H^1-norm for piecewise polynomials in V_0^h. By Lemma 5.1 in [42], we have the following standard V^h-ellipticity result for a domain on which elliptic regularity holds, e.g., $\Omega = [-L, L]^d$.

Lemma C.2. *Let* $\|v_h\|_X = \sqrt{\langle \nabla v_h, \nabla v_h \rangle + \alpha \langle v_h, v_h \rangle} = \|\mathbf{v}\|_X$. *Let* Ω *be a domain with elliptic regularity. There is a constant* $D_4 > 0$ *independent of mesh size* h *s.t.*

$$(C.2) \qquad \forall v_h \in V_0^h, \quad \frac{1}{D_4} \|v_h\|_X \le \|v_h\|_{H^1(\Omega)} \le D_4 \|v_h\|_X.$$

C.2. **Proofs of Lemma 5.4 and Lemma 5.5.**

Proof of Lemma 5.4. Since $\langle \mathbf{u}, \mathbf{v} \rangle_h = 0$, we have that $\|\mathbf{u} + \mathbf{v}\|_2^2 = \|\mathbf{u}\|_2^2 + \|\mathbf{v}\|_2^2 = 1 + \|\mathbf{v}\|_2^2$. This implies that

$$R_h(\mathbf{u} + \mathbf{v}) - (\mathbf{u} + \mathbf{v}) = \left(\frac{1}{\|\mathbf{u} + \mathbf{v}\|_2} - 1 \right)(\mathbf{u} + \mathbf{v}) = \left(\left(1 + \|\mathbf{v}\|_2^2 \right)^{-\frac{1}{2}} - 1 \right)(\mathbf{u} + \mathbf{v}).$$

Then (5.1) follows from the elementary inequality $1 - \frac{x}{2} \le (1 + x)^{-1/2} \le 1$, $x \ge 0$. $\qquad\square$

Proof of Lemma 5.5. (i) $\|\mathbf{u}\|_X^2 = \mathbf{u}^\top (\mathbb{S} + \alpha \mathbb{M})\mathbf{u} \ge \alpha \mathbf{u}^\top \mathbb{M} \mathbf{u} = \alpha \|\mathbf{u}\|_2^2$.

(ii) By Cauchy-Schwartz inequality for the inner product $\langle \cdot, \cdot \rangle_h$ and (i), we get

$$\left\| (-\Delta_h + \alpha \mathbb{I})^{-1} \mathbf{u} \right\|_X^2 = \left\langle (-\Delta_h + \alpha \mathbb{I})^{-1} \mathbf{u}, \mathbf{u} \right\rangle_h \le \left\| (-\Delta_h + \alpha \mathbb{I})^{-1} \mathbf{u} \right\|_2 \|\mathbf{u}\|_2 \le \frac{1}{\sqrt{\alpha}} \left\| (-\Delta_h + \alpha \mathbb{I})^{-1} \mathbf{u} \right\|_X \|\mathbf{u}\|_2.$$

(iii) Since $\Pi(\mathbf{u}^2)$ coincides with u_h^2 at the quadrature nodes, by (C.1) for $p = 2$,

$$\|\mathbf{u}^2\|_2 = \sqrt{\langle \mathbf{u}^2, \mathbf{u}^2 \rangle_h} = \sqrt{\langle \Pi(\mathbf{u}^2), \Pi(\mathbf{u}^2) \rangle} = \sqrt{\langle u_h^2, u_h^2 \rangle} \le D_2^2 \|u_h\|_{L^4(\Omega)}^2.$$

With the Sobolev embedding $H^1(\Omega) \subset L^4(\Omega)$ for dimension $d \le 3$, and (C.2), we get

$$\|u_h\|_{L^4(\Omega)}^2 \le D^2 \|u_h\|_{H^1(\Omega)}^2 \le D^2 D_4^2 \|\mathbf{u}\|_X^2,$$

where $D > 0$ is the constant associated with the embedding $H^1(\Omega) \subset L^4(\Omega)$.

(iv) The proof is almost identical to that of (iii), which uses (C.1) for $p = 3$, (C.2), and the Sobolev embedding $H^1(\Omega) \subset L^6(\Omega)$ for dimension $d = 1, 2, 3$. $\qquad\square$

APPENDIX D. DEFERRED PROOFS FOR SECTION 5.2

Proof of Lemma 5.11. Since $A_{\mathbf{u}^*}$ is self-adjoint w.r.t. $\langle \cdot, \cdot \rangle_h$, it has orthonormal eigenvectors. Let $\mathbf{u}_\parallel = \langle \mathbf{u}, \mathbf{u}^* \rangle_h \mathbf{u}^*$ be the orthogonal projection of $\mathbf{u}$ onto the subspace spanned by $\mathbf{u}^*$. Let $\mathbf{u}_\perp = \mathbf{u} - \mathbf{u}_\parallel$, then $\langle \mathbf{u}_\parallel, \mathbf{u}_\perp \rangle_h = 0$ and $\|\mathbf{u}_\parallel\|_2^2 + \|\mathbf{u}_\perp\|_2^2 =$

1. Thus,

$$
\begin{aligned}
\mathbf{E}_h(\mathbf{u}) - \mathbf{E}_h(\mathbf{u}^*) &= \frac{1}{2}\langle -\Delta_h \mathbf{u}, \mathbf{u}\rangle_h + \frac{1}{2}\langle \mathbb{V}\mathbf{u}, \mathbf{u}\rangle_h + \frac{\beta}{4}\langle \mathbf{u}^2, \mathbf{u}^2\rangle_h \\
&\quad - \frac{1}{2}\langle -\Delta_h \mathbf{u}^*, \mathbf{u}^*\rangle_h - \frac{1}{2}\langle \mathbb{V}\mathbf{u}^*, \mathbf{u}^*\rangle_h - \frac{\beta}{4}\langle (\mathbf{u}^*)^2, (\mathbf{u}^*)^2\rangle_h \\
&\geq \frac{1}{2}\langle -\Delta_h \mathbf{u}, \mathbf{u}\rangle_h + \frac{1}{2}\langle \mathbb{V}\mathbf{u}, \mathbf{u}\rangle_h + \frac{\beta}{2}\langle (\mathbf{u}^*)^2, \mathbf{u}^2\rangle_h \\
&\quad - \frac{1}{2}\langle -\Delta_h \mathbf{u}^*, \mathbf{u}^*\rangle_h - \frac{1}{2}\langle \mathbb{V}\mathbf{u}^*, \mathbf{u}^*\rangle_h - \frac{\beta}{2}\langle (\mathbf{u}^*)^2, (\mathbf{u}^*)^2\rangle_h \\
&= \frac{1}{2}\langle A_{\mathbf{u}^*}\mathbf{u}, \mathbf{u}\rangle_h - \frac{1}{2}\langle A_{\mathbf{u}^*}\mathbf{u}^*, \mathbf{u}^*\rangle_h = \frac{1}{2}\langle A_{\mathbf{u}^*}\mathbf{u}_\|, \mathbf{u}_\|\rangle_h + \frac{1}{2}\langle A_{\mathbf{u}^*}\mathbf{u}_\perp, \mathbf{u}_\perp\rangle_h - \frac{\lambda_h^0}{2} \\
&\geq \frac{\lambda_h^0}{2}\left\|\mathbf{u}_\|\right\|_2^2 + \frac{\lambda_h^1}{2}\left\|\mathbf{u}_\perp\right\|_2^2 - \frac{\lambda_0}{2} = \frac{\lambda_h^1 - \lambda_h^0}{2}\left\|\mathbf{u}_\perp\right\|_2^2.
\end{aligned}
$$

With the following fact, (5.5) follows from the estimate above:

$$
\begin{aligned}
\left\|\mathbf{u}_\perp\right\|_2^2 = 1 - \left\|\mathbf{u}_\|\right\|_2^2 = 1 - |\langle \mathbf{u}, \mathbf{u}^*\rangle_h|^2 &= 1 - \frac{1}{4}\left(2 - \left\|\mathbf{u} - \mathbf{u}^*\right\|_2^2\right)^2 \\
&= \left\|\mathbf{u} - \mathbf{u}^*\right\|_2^2 - \frac{1}{4}\left\|\mathbf{u} - \mathbf{u}^*\right\|_2^4.
\end{aligned}
$$

$\square$

Lemma D.1. *There is some constant $L_1 > 0$ independent of the mesh size h s.t.*

$$
\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^N, \quad \left\|(-\Delta_h + \alpha\mathbb{I})^{-1}(\mathbf{u}^3 - \mathbf{v}^3)\right\|_X \leq L_1 \left\|\mathbf{u} - \mathbf{v}\right\|_X \left(\left\|\mathbf{u}\right\|_X^4 + \left\|\mathbf{v}\right\|_X^4\right).
$$

Proof. Using Lemma 5.5 (ii) and (iv), we can compute that

$$
\begin{aligned}
\left\|(-\Delta_h + \alpha\mathbb{I})^{-1}(\mathbf{u}^3 - \mathbf{v}^3)\right\|_X^2 &\leq \frac{1}{\alpha}\left\|\mathbf{u}^3 - \mathbf{v}^3\right\|_2^2 = \frac{1}{\alpha}\sum_i w_i(u_i^3 - v_i^3)^2 \\
&= \frac{1}{\alpha}\sum_i w_i(u_i - v_i)^2(u_i^2 + u_iv_i + v_i^2)^2 \\
&\leq \frac{8}{\alpha}\sum_i w_i(u_i - v_i)^2 u_i^4 + \frac{8}{\alpha}\sum_i w_i(u_i - v_i)^2 v_i^4 \\
&\leq \frac{8}{\alpha}\left(\sum_i w_i(u_i-v_i)^6\right)^{1/3}\left(\sum_i w_i u_i^6\right)^{2/3} + \frac{8}{\alpha}\left(\sum_i w_i(u_i-v_i)^6\right)^{1/3}\left(\sum_i w_i v_i^6\right)^{2/3} \\
&= \frac{8}{\alpha}\left\|(\mathbf{u} - \mathbf{v})^3\right\|_2^{2/3}\left(\left\|\mathbf{u}^3\right\|_2^{4/3} + \left\|\mathbf{v}^3\right\|_2^{4/3}\right) \\
&\leq \frac{8C_2^2}{\alpha}\left\|\mathbf{u} - \mathbf{v}\right\|_X^2\left(\left\|\mathbf{u}\right\|_X^4 + \left\|\mathbf{v}\right\|_X^4\right).
\end{aligned}
$$

$\square$

Lemma D.2. *For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$ with $\|\mathbf{u}\|_X \leq 2\|\mathbf{u}^*\|_X$ and $\|\mathbf{v}\|_X \leq 2\|\mathbf{u}^*\|_X$, it holds for some constant $L_2 > 0$ depending on $\|\mathbf{u}^*\|_X$ and independent of the mesh size h that $\|\nabla_X \mathbf{E}_h(\mathbf{u}) - \nabla_X \mathbf{E}_h(\mathbf{v})\|_X \leq L_2 \|\mathbf{u} - \mathbf{v}\|_X$.*

Proof. Recall that $\nabla_X \mathbf{E}_h(\mathbf{u}) = \mathbf{u} + (-\Delta_h + \alpha\mathbb{I})^{-1}(\mathbb{V} - \alpha\mathbb{I} + \beta\mathrm{diag}(\mathbf{u})^2)\mathbf{u}$. Using Lemma D.1 and Lemma 5.5, we can conclude that

$$
\begin{aligned}
\|\nabla_X &\mathbf{E}_h(\mathbf{u}) - \nabla_X \mathbf{E}_h(\mathbf{v})\|_X \\
&\leq \|\mathbf{u}-\mathbf{v}\|_X + \left\|(-\Delta_h + \alpha\mathbb{I})^{-1}(\mathbb{V}-\alpha\mathbb{I})(\mathbf{u}-\mathbf{v})\right\|_X + \beta\left\|(-\Delta_h+\alpha\mathbb{I})^{-1}(\mathbf{u}^3-\mathbf{v}^3)\right\|_X \\
&\leq \|\mathbf{u}-\mathbf{v}\|_X + \frac{V_{\alpha,\max}}{\sqrt{\alpha}}\|\mathbf{u}-\mathbf{v}\|_2 + \beta L_1\|\mathbf{u}-\mathbf{v}\|_X\left(\|\mathbf{u}\|_X^4 + \|\mathbf{v}\|_X^4\right) \\
&\leq \left(1 + \frac{V_{\alpha,\max}}{\alpha} + 32\beta L_1\|\mathbf{u}^*\|_X^4\right)\|\mathbf{u}-\mathbf{v}\|_X.
\end{aligned}
$$

$\square$

Proof of Lemma 5.12. Recall $\nabla_X^{\mathcal{R}}\mathbf{E}_h(\mathbf{u}) = \nabla_X\mathbf{E}_h(\mathbf{u}) - \frac{\langle\mathbf{u},\nabla_X\mathbf{E}_h(\mathbf{u})\rangle_h}{\langle\mathbb{G}_X\mathbf{u},\mathbf{u}\rangle_h}\mathbb{G}_X\mathbf{u}$. Set

$$
\gamma = \frac{\langle\mathbf{u},\nabla_X\mathbf{E}_h(\mathbf{u})\rangle_h}{\langle\mathbb{G}_X\mathbf{u},\mathbf{u}\rangle_h} \quad\text{and}\quad \gamma^* = \frac{\langle\mathbf{u}^*,\nabla_X\mathbf{E}_h(\mathbf{u}^*)\rangle_h}{\langle\mathbb{G}_X\mathbf{u}^*,\mathbf{u}^*\rangle_h} = \frac{\langle\mathbf{u}^*,\mathbb{G}_X A_{\mathbf{u}^*}\mathbf{u}^*\rangle_h}{\langle\mathbb{G}_X\mathbf{u}^*,\mathbf{u}^*\rangle_h} = \lambda_h^0.
$$

Suppose that $\|\mathbf{u}-\mathbf{u}^*\|_X \leq \|\mathbf{u}^*\|_X$, then $\|\mathbf{u}\|_X \leq 2\|\mathbf{u}^*\|_X$. With Lemma 5.5 and Lemma D.2, it holds that

$$
\begin{aligned}
\|\nabla_X^{\mathcal{R}}&\mathbf{E}_h(\mathbf{u})\|_X \\
&= \left\|\nabla_X^{\mathcal{R}}\mathbf{E}_h(\mathbf{u}) - \nabla_X^{\mathcal{R}}\mathbf{E}_h(\mathbf{u}^*)\right\|_X \\
&= \|\nabla_X\mathbf{E}_h(\mathbf{u}) - \gamma\mathbb{G}_X\mathbf{u} - \nabla_X\mathbf{E}_h(\mathbf{u}^*) + \gamma^*\mathbb{G}_X\mathbf{u}^*\|_X \\
&\leq \|\nabla_X\mathbf{E}_h(\mathbf{u}) - \nabla_X\mathbf{E}_h(\mathbf{u}^*)\|_X + |\gamma-\gamma^*|\cdot\|\mathbb{G}_X\mathbf{u}\|_X + |\gamma^*|\cdot\|\mathbb{G}_X\mathbf{u}-\mathbb{G}_X\mathbf{u}^*\|_X \\
&\leq L_2\|\mathbf{u}-\mathbf{u}^*\|_X + \frac{\|\mathbf{u}^*\|_X}{\alpha}|\gamma-\gamma^*| + \frac{|\gamma^*|}{\alpha}\|\mathbf{u}-\mathbf{u}^*\|_X.
\end{aligned}
$$

Thus it suffices to estimate $|\gamma-\gamma^*|$. Set

$$
\begin{aligned}
A &= \langle\mathbf{u},\nabla_X\mathbf{E}_h(\mathbf{u})\rangle_h, \quad A^* = \langle\mathbf{u}^*,\nabla_X\mathbf{E}_h(\mathbf{u}^*)\rangle_h, \\
B &= \langle\mathbb{G}_X\mathbf{u},\mathbf{u}\rangle_h, \quad B^* = \langle\mathbb{G}_X\mathbf{u}^*,\mathbf{u}^*\rangle_h.
\end{aligned}
$$

By Lemma 5.5, Lemma D.2, and $\|\mathbf{u}\|_2 = \|\mathbf{u}^*\|_2 = 1$, we have

$$
\begin{aligned}
|A-A^*| &\leq \|\mathbf{u}-\mathbf{u}^*\|_2\|\nabla_X\mathbf{E}_h(\mathbf{u}^*)\|_2 + \|\mathbf{u}\|_2\|\nabla_X\mathbf{E}_h(\mathbf{u})-\nabla_X\mathbf{E}_h(\mathbf{u}^*)\|_2 \\
&\leq \frac{\|\nabla_X\mathbf{E}_h(\mathbf{u}^*)\|_X}{\alpha}\|\mathbf{u}-\mathbf{u}^*\|_X + \frac{L_2}{\sqrt{\alpha}}\|\mathbf{u}\|_X\|\mathbf{u}-\mathbf{u}^*\|_X,
\end{aligned}
$$

$$
|B-B^*| \leq \|\mathbf{u}-\mathbf{u}^*\|_2\|\mathbb{G}_X\mathbf{u}\|_2 + \|\mathbf{u}^*\|_2\|\mathbb{G}_X\mathbf{u}-\mathbb{G}_X\mathbf{u}^*\|_2 \leq \frac{2}{\alpha^{3/2}}\|\mathbf{u}-\mathbf{u}^*\|_X,
$$

$$
|\gamma-\gamma^*| = \left|\frac{A}{B} - \frac{A^*}{B^*}\right| \leq \frac{|A-A^*|\cdot|B^*| + |A^*|\cdot|B-B^*|}{|BB^*|} \leq L_\gamma\|\mathbf{u}-\mathbf{u}^*\|_X,
$$

for small $\|\mathbf{u}-\mathbf{u}^*\|_X$ and some constant $L_\gamma > 0$, which completes the proof. $\square$

Appendix E. Positive eigengap independent of the mesh size

In this section, we prove that Assumption 5.8 holds with a positive eigengap independent of the mesh size for the P^1 finite element method with quadrature on an unstructured simplicial mesh with shape regularity.

Definition E.1 (Shape regular meshes). A simplicial mesh with h_K denoting the size of each simplex K is shape regular if there exists some constant $\sigma > 0$ independent of the mesh size $h = \sup_K h_K$, such that $\frac{h_K}{\rho_K} \leq \sigma$ holds for all K, where ρ_K is the diameter of the largest sphere contained in K.

E.1. **Technical lemmas.**

Lemma E.2. *Suppose that $V \in C^1(\bar{\Omega})$. There exists constants $E_1, E_2, E_3 > 0$ independent of the mesh size, so that for any $h > 0$ and $\mathbf{u} \in \mathbb{R}^N$, the following holds for P^1 finite element method with quadrature on an unstructured simplicial mesh with shape regularity:*

(i) $0 \leq \|\mathbf{u}\|_2^2 - \|\Pi\mathbf{u}\|_{L^2(\Omega)}^2 \leq E_1 h \langle -\Delta_h \mathbf{u}, \mathbf{u} \rangle_h.$

(ii) $\left| \langle \mathbf{u}, \mathbb{V}\mathbf{u} \rangle_h - \int_\Omega V|\Pi\mathbf{u}|^2 \right| \leq E_2 h \langle -\Delta_h \mathbf{u}, \mathbf{u} \rangle_h + E_3 h \|\mathbf{u}\|_2^2.$

(iii) $\langle -\Delta_h \mathbf{u}, \mathbf{u} \rangle_h = |\Pi\mathbf{u}|_{H_0^1(\Omega)}^2.$

(iv) $\langle \mathbf{u}^2, \mathbf{u}^2 \rangle_h \geq \|\Pi\mathbf{u}\|_{L^4(\Omega)}^4.$

Proof. It suffices to prove the results on each simplex K of the mesh since all of them are additive. Consider a single simplex K. For convenience, by abusing notation, we let all vertices of K be indexed as $\{\boldsymbol{x}_i, i = 1, \cdots, d+1\}$ and the P^1 interpolation polynomial $(\Pi\mathbf{u})(\mathbf{x}) = a + \mathbf{b}^\top \mathbf{x}$ on K satisfying $(\Pi\mathbf{u})(\mathbf{x}_i) = u_i$. Define the barycentric coordinates as $\sum_{i=1}^{d+1} t_i = 1, t_i \geq 0$ for the simplex K, i.e., $\boldsymbol{x} = \sum_{i=1}^{d+1} t_i \boldsymbol{x}_i, \forall \boldsymbol{x} \in K$. It is straightforward to verify that the polynomial $[\Pi\mathbf{u}(\boldsymbol{x})]^2$ is a convex function of $\boldsymbol{x}$, thus we have Jensen's inequality

$$(\Pi\mathbf{u})^2 \left(\sum_{i=1}^{d+1} t_i \mathbf{x}_i \right) \leq \sum_{i=1}^{d+1} t_i (\Pi\mathbf{u})^2(\mathbf{x}_i).$$

Consider a barycentric simplex $\Delta_I = \{(t_1, \cdots, t_{d+1}) \in [0,1]^{d+1} : \sum_{i \in I} t_i = 1\} \subset \mathbb{R}^{d+1}$. By regarding Δ_I as an embedded manifold in $\mathbb{R}^{d+1}$, its volume $|\Delta_I|$ can be written as an integral of constant 1:

$$|\Delta_I| = \int_{\Delta_I} 1 dV = \int_{\Delta_I} \left(\sum_{i=1}^{d+1} t_i \right) dV = (d+1) \int_{\Delta_I} t_j dV, \quad \text{for any fixed } j,$$

where dV is the volume form of the manifold and in the last step we have used the symmetry of integrating each t_j on the simplex Δ_I. Since $\frac{1}{d+1} = \frac{1}{|\Delta_I|} \int_{\Delta_I} (t_j) \, dV$, $\forall j$, we have

(E.1)
$$\frac{|K|}{d+1} \sum_{i=1}^{d+1} u_i^2 = \frac{|K|}{d+1} \sum_{i=1}^{d+1} (\Pi\mathbf{u})^2(\mathbf{x}_i) = \frac{|K|}{|\Delta_I|} \int_{\Delta_I} \sum_{i=1}^{d+1} t_i (\Pi\mathbf{u})^2(\mathbf{x}_i) dV$$

$$\geq \frac{|K|}{|\Delta_I|} \int_{\Delta_I} (\Pi\mathbf{u})^2 \left(\sum_{i=1}^{d+1} t_i \mathbf{x}_i \right) dV = \int_K (\Pi\mathbf{u})^2(\mathbf{x}) d\mathbf{x}.$$

Summing the inequality above over all simplexes in the mesh, the first inequality in (i) is proven. Note that the quadrature $\frac{|K|}{d+1} \sum_{i=1}^{d+1} f(\boldsymbol{x}_i) = \frac{|K|}{d+1} \sum_{\mathbf{x}_i \in K} f(\boldsymbol{x}_i)$ is equal to $\int_K f(\boldsymbol{x}) d\boldsymbol{x}$ for any linear function $f(\boldsymbol{x})$. Let $E_1 = 2 \sup_{\mathbf{x} \in \Omega} \|\mathbf{x}\|$, then we

have

(E.2)
$$\frac{|K|}{d+1} \sum_{\mathbf{x}_i \in K} (a + \mathbf{b}^\top \mathbf{x}_i)^2 - \int_K (a + \mathbf{b}^\top \mathbf{x})^2 \mathrm{d}\mathbf{x} = \frac{|K|}{d+1} \sum_{\mathbf{x}_i \in K} (\mathbf{b}^\top \mathbf{x}_i)^2 - \int_K (\mathbf{b}^\top \mathbf{x})^2 \mathrm{d}\mathbf{x}$$
$$\leq |K| \sup_{\mathbf{x}, \mathbf{x}' \in K} \left((\mathbf{b}^\top \mathbf{x})^2 - (\mathbf{b}^\top \mathbf{x}')^2 \right) = |K| \sup_{\mathbf{x}, \mathbf{x}' \in K} \left(\mathbf{b}^\top \mathbf{x} - \mathbf{b}^\top \mathbf{x}' \right) \left(\mathbf{b}^\top \mathbf{x} + \mathbf{b}^\top \mathbf{x}' \right)$$
$$\leq E_1 h |K| \|\mathbf{b}\|^2.$$

With notation in (2.11), the discrete integration by parts (2.10) can be also be written as

(E.3)
$$\langle -\Delta_h \mathbf{u}, \mathbf{u} \rangle_h = \langle \nabla u_h, \nabla u_h \rangle = \langle \nabla \Pi \mathbf{u}(\boldsymbol{x}), \nabla \Pi \mathbf{u}(\boldsymbol{x}) \rangle.$$

Notice that $\nabla \Pi \mathbf{u}(\mathbf{x}) = \mathbf{b}$ is a constant over K, thus after summing over K, the term $|K| \|\mathbf{b}\|^2$ becomes $\langle \nabla \Pi \mathbf{u}(\boldsymbol{x}), \nabla \Pi \mathbf{u}(\boldsymbol{x}) \rangle = \langle -\Delta_h \mathbf{u}, \mathbf{u} \rangle_h$. By the same notation, we also have

(E.4)
$$\sum_K \frac{|K|}{d+1} \sum_{\mathbf{x}_i \in K} (a + \mathbf{b}^\top \mathbf{x}_i)^2 = \langle \Pi \mathbf{u}(\boldsymbol{x}), \Pi \mathbf{u}(\boldsymbol{x}) \rangle = \langle \mathbf{u}, \mathbf{u} \rangle_h,$$
$$\sum_K \int_K (a + \mathbf{b}^\top \mathbf{x})^2 \mathrm{d}\mathbf{x} = \|\Pi \mathbf{u}\|_{L^2(\Omega)}^2,$$

which proves the second inequality of (i).

Notice that (E.1) can also be written as

(E.5)
$$\int_K (a + \mathbf{b}^\top \mathbf{x})^2 \mathrm{d}\mathbf{x} \leq \frac{|K|}{d+1} \sum_{\mathbf{x}_i \in K} (a + \mathbf{b}^\top \mathbf{x}_i)^2.$$

For proving (ii), with some fixed $\tilde{\mathbf{x}} \in K$, we have

$$\left| \frac{|K|}{d+1} \sum_{\mathbf{x}_i \in K} V(\mathbf{x}_i)(a + \mathbf{b}^\top \mathbf{x}_i)^2 - \int_K V(\mathbf{x})(a + \mathbf{b}^\top \mathbf{x})^2 \mathrm{d}\mathbf{x} \right|$$
$$\leq |V(\tilde{\mathbf{x}})| \cdot \left| \frac{|K|}{d+1} \sum_{\mathbf{x}_i \in K} (a + \mathbf{b}^\top \mathbf{x}_i)^2 - \int_K (a + \mathbf{b}^\top \mathbf{x})^2 \mathrm{d}\mathbf{x} \right|$$
$$+ \frac{|K|}{d+1} \sum_{\mathbf{x}_i \in K} |V(\mathbf{x}_i) - V(\tilde{\mathbf{x}})|(a + \mathbf{b}^\top \mathbf{x}_i)^2 + \int_K |V(\mathbf{x}) - V(\tilde{\mathbf{x}})|(a + \mathbf{b}^\top \mathbf{x})^2 \mathrm{d}\mathbf{x}$$
$$\leq \|V\|_{L^\infty(\bar{\Omega})} \left| \frac{|K|}{d+1} \sum_{\mathbf{x}_i \in K} (a + \mathbf{b}^\top \mathbf{x}_i)^2 - \int_K (a + \mathbf{b}^\top \mathbf{x})^2 \mathrm{d}\mathbf{x} \right|$$
$$+ h_K \|\nabla V\|_{L^\infty(\bar{\Omega})} \left(\frac{|K|}{d+1} \sum_{\mathbf{x}_i \in K} (a + \mathbf{b}^\top \mathbf{x}_i)^2 + \int_K (a + \mathbf{b}^\top \mathbf{x})^2 \mathrm{d}\mathbf{x} \right)$$
$$\leq \|V\|_{L^\infty(\bar{\Omega})} E_1 h |K| \|\mathbf{b}\|^2 + 2 h_K \|\nabla V\|_{L^\infty(\bar{\Omega})} \frac{|K|}{d+1} \sum_{\mathbf{x}_i \in K} (a + \mathbf{b}^\top \mathbf{x}_i)^2,$$

where we have used (E.2) and (E.5) in the last step. By summing over K, we have proven (ii). Since $\nabla \Pi \mathbf{u}$ is a constant over K, the quadrature for $(\nabla \Pi \mathbf{u})^2$ is exact, i.e., $\langle \nabla \Pi \mathbf{u}(\boldsymbol{x}), \nabla \Pi \mathbf{u}(\boldsymbol{x}) \rangle = \|\nabla \Pi \mathbf{u}\|_{L^2(\Omega)}^2$, thus by (E.3) the result (iii) is true.

Finally, with a similar argument as in (E.1), the result (iv) is also a consequence of Jensen's inequality since $[\Pi\mathbf{u}(\boldsymbol{x})]^4$ is a convex function on K. $\qquad\square$

Lemma E.3. *Consider P^1 finite element method with quadrature on an unstructured simplicial mesh with shape regularity in dimension $d = 1, 2, 3$. For any $h > 0$ and $u \in H^2(\Omega) \cap H_0^1(\Omega)$, it holds for some constant $E_4 > 0$ independent of the mesh size that $\left| |u|_{H_0^1(\Omega)}^2 - \langle -\Delta_h \mathbf{u}, \mathbf{u} \rangle_h \right| \leq E_4 h \|u\|_{H^2(\Omega)}^2$, where $\mathbf{u} \in \mathbb{R}^N$ is the vector of the values of u at quadrature points.*

Remark E.4. Lemma E.2 holds for any dimension d, but Lemma E.3 holds only for dimension $d \leq 3$ since we need the point values $u(\boldsymbol{x}_i)$ being well defined, ensured by the Sobolev embedding $u \in H^2(\Omega) \cap H_0^1(\Omega) \subset C(\Omega)$, which holds only in dimension $d = 1, 2, 3$.

Proof. By Lemma E.2 (iii), we have

$$\left| |u|_{H_0^1(\Omega)}^2 - \langle -\Delta_h \mathbf{u}, \mathbf{u} \rangle_h \right| = \left| |u|_{H_0^1(\Omega)}^2 - |\Pi\mathbf{u}|_{H_0^1(\Omega)}^2 \right|$$

$$\leq |u - \Pi\mathbf{u}|_{H_0^1(\Omega)} |u + \Pi\mathbf{u}|_{H_0^1(\Omega)} \leq |u - \Pi\mathbf{u}|_{H_0^1(\Omega)} \left(2|u|_{H_0^1(\Omega)} + |u - \Pi\mathbf{u}|_{H_0^1(\Omega)} \right).$$

We can conclude the proof after combining the estimates above and the standard interpolation estimate $|u - \Pi\mathbf{u}|_{H_0^1(\Omega)} = O(h) \cdot \|u\|_{H^2(\Omega)}$, see [18]. $\qquad\square$

Lemma E.5. *Under the same assumptions as in Lemma E.2, let $\bar{\Pi}\mathbf{u}$ be the piecewise constant interpolation of $\mathbf{u}$, i.e., $(\bar{\Pi}\mathbf{u})(\mathbf{x}) = \frac{1}{|K|} \int_K (\Pi\mathbf{u})(\mathbf{x}') d\mathbf{x}'$ on any simplex K. There is some constant $E_5 > 0$ independent of the mesh size s.t. $\|\Pi\mathbf{u} - \bar{\Pi}\mathbf{u}\|_{L^2(\Omega)}^2 \leq E_5 h^2 \langle -\Delta_h \mathbf{u}, \mathbf{u} \rangle_h$.*

Proof. Consider a simplex K on which we assume $(\Pi\mathbf{u})(\mathbf{x}) = a + \mathbf{b}^\top \mathbf{x}$ and $(\bar{\Pi}\mathbf{u})(\mathbf{x}) = a + \frac{\mathbf{b}^\top}{3} \sum_{\mathbf{x}_i \in K} \mathbf{x}_i$. Then we have

$$\int_K ((\Pi\mathbf{u})(\mathbf{x}) - (\bar{\Pi}\mathbf{u})(\mathbf{x}))^2 = \int_K \left(\mathbf{b}^\top \mathbf{x} - \frac{\mathbf{b}^\top}{3} \sum_{\mathbf{x}_i \in K} \mathbf{x}_i \right)^2 d\mathbf{x} \leq h_K^2 |K| \|\mathbf{b}\|^2 \leq h^2 |K| \|\mathbf{b}\|^2.$$

By the same arguments as in the proof for Lemma E.2, the result holds with $E_5 = 1$. $\qquad\square$

E.2. **Proof of consistency and eigengap.**

Theorem E.6. *Consider P^1 finite element method with quadrature on an unstructured simplicial mesh with shape regularity in dimension $d = 1, 2, 3$. Let u^* and λ^* be the ground state and the corresponding eigenvalue of the continuous energy E. Let u_h^* and λ_h^* be the ground state and the corresponding eigenvalue of the discrete energy E_h. Suppose that $u^* \in H^2(\Omega) \cap H_0^1(\Omega)$ and that $V \in C^1(\bar{\Omega})$. Then $u_h^* \to u^*$ in $H_0^1(\Omega)$ and $\lambda_h^* \to \lambda^*$ as $h \to 0$.*

Proof. By our previous notation, $\mathbf{u}^* \in \mathbb{R}^N$ is the vector collecting the values of u_h^* at quadrature points and $\Pi(\mathbf{u}^*) = u_h^*$. For convenience, define $\mathbf{u}^{exact} \in \mathbb{R}^N$ as the vector representing the values of u^* at quadrature points. Due to the embedding $H^2(\Omega) \hookrightarrow C^0(\Omega)$ for $d = 1, 2, 3$ and the convergence of Riemannian integral for continuous functions, one has $\langle \mathbf{u}^{exact}, \mathbf{u}^{exact} \rangle_h = \|u^*\|_{L^2(\Omega)}^2 + o(1) = 1 + o(1)$, $\langle \mathbf{u}^{exact}, \mathbb{V}\mathbf{u}^{exact} \rangle_h = \int V|u^*|^2 + o(1)$, and $\langle ((\mathbf{u}^{exact})^2, (\mathbf{u}^{exact})^2) \rangle_h = \int |u^*|^4 + o(1)$.

Furthermore, thanks to Lemma E.3, we have $\mathbf{E}_h(\mathbf{u}^{exact}) = E(u^*) + o(1)$. Then it holds that

$$\text{(E.6)} \quad \mathbf{E}_h(\mathbf{u}^*) \le \mathbf{E}_h\left(\frac{\mathbf{u}^{exact}}{\langle \mathbf{u}^{exact}, \mathbf{u}^{exact}\rangle_h}\right) = \mathbf{E}_h(\mathbf{u}^{exact}) + o(1) = E(u^*) + o(1),$$

which implies that $\langle -\Delta_h \mathbf{u}^*, \mathbf{u}^*\rangle_h$ is bounded as $h \to 0$.

By Lemma E.2 (i) and boundedness of $\langle -\Delta_h \mathbf{u}^*, \mathbf{u}^*\rangle_h$, one has $\|u_h^*\|_{L^2(\Omega)} = 1 + o(1)$. Therefore, it follows from Lemma E.2 (ii)–(iv) that

$$\text{(E.7)} \quad E_h(u_h^*) = \mathbf{E}_h(\mathbf{u}^*) \ge E(u_h^*) + o(1) = E\left(\frac{u_h^*}{\|u_h^*\|_{L^2(\Omega)}}\right) + o(1) \ge E(u^*) + o(1).$$

Then we can conclude from (E.6) and (E.7) that $E(u^*) = E\left(\frac{u_h^*}{\|u_h^*\|_{L^2(\Omega)}}\right) + o(1)$.

By compact embeddings $H_0^1(\Omega) \subset\subset L^2(\Omega)$ and $H_0^1(\Omega) \subset\subset L^4(\Omega)$, there exists a subsequence of the bounded sequence $\left\{\frac{u_h^*}{\|u_h^*\|_{L^2(\Omega)}}\right\}$ converging weakly in $H_0^1(\Omega)$ and strongly in $L^2(\Omega)$ and $L^4(\Omega)$ to some function in $\in H_0^1(\Omega)$. Without loss of generality, we assume that the whole sequence converges:

$$\frac{u_h^*}{\|u_h^*\|_{L^2(\Omega)}} \to u_0, \quad \text{weakly in } H_0^1(\Omega) \text{ and strongly in } L^2(\Omega) \text{ and } L^4(\Omega).$$

Since the weak convergence guarantees that

$$\|u^0\|_{H_0^1(\Omega)} \le \liminf_{h\to 0} \left\|\frac{u_h^*}{\|u_h^*\|_{L^2(\Omega)}}\right\|_{H_0^1(\Omega)},$$

$$E(u^*) \le E(u^0) \le \lim_{h\to 0} E\left(\frac{u_h^*}{\|u_h^*\|_{L^2(\Omega)}}\right) = E(u^*),$$

which implies that $E(u^0) = E(u^*)$ and hence that $u^0 = u^*$ by the uniqueness of positive ground state. Then we know that $\frac{u_h^*}{\|u_h^*\|_{L^2(\Omega)}} \to u^*$ weakly in $H_0^1(\Omega)$ and $\left\|\frac{u_h^*}{\|u_h^*\|_{L^2(\Omega)}}\right\|_{H_0^1(\Omega)} \to \|u^*\|_{H_0^1(\Omega)}$. Note that the weak convergence and convergence of norm imply the strong convergence. Thus $\frac{u_h^*}{\|u_h^*\|_{L^2(\Omega)}} \to u^*$ and $u_h^* \to u^*$ strongly in $H_0^1(\Omega)$.

In the discussion above, we have shown that every subsequence of $\{u_h^*\}$ has a subsequence that converges strongly in $H_0^1(\Omega)$ to u^*. Therefore, the whole sequence $\{u_h^*\}$ converges strongly in $H_0^1(\Omega)$ to u^*. This immediately implies that $\lambda_h^* \to \lambda^*$. $\qquad\square$

Assumption E.7. We assume that the multiplicity of the eigenvalue λ^* to the following problem is one: $-\Delta u + Vu + \beta|u^*|^2 u = \lambda u, \quad u = 0$ on $\partial\Omega$.

We remark that the smoothness of u^* depends on the smoothness of V [44], and the assumption above can be proven with suitable smoothness assumptions on V and u^*, e.g., the smallest eigenvalue of an elliptic operator $-\Delta + V + \beta|u^*|^2$ is simple due to the positivity and the maximum principle of the operator $(-\Delta + V + \beta|u^*|^2)^{-1}$, implied by the Krein-Rutman Theorem (the extension of Perron-Frobenius Theorem from positive matrices to positive compact linear operators),

see [25, Chapter 1]. The spectrum, including simplicity of the smallest eigenvalue, of Schrödinger operators like $-\Delta + V + \beta|u^*|^2$, was also thoroughly studied in [48, Section XIII.12].

Theorem E.8. *Suppose that Assumption* E.7 *and assumptions made in Theorem* E.6 *hold. Let* $(\lambda_h^1, \mathbf{u}^1)$ *with* $\langle \mathbf{u}^1, \mathbf{u}^1 \rangle_h = 1$ *be the second eigenpair of the linearized problem* $-\Delta_h \mathbf{u} + \mathbb{V}\mathbf{u} + \beta(\mathbf{u}^*)^2\mathbf{u} = \lambda_h \mathbf{u}$. *Then* $\liminf_{h\to 0} \lambda_h^1 - \lambda_h^* > 0$.

We remark that λ_h^* in Theorem E.8 is exactly λ_h^0 in Assumption 5.8. In this sense Theorem E.8 validates Assumption 5.8 with a positive lower bound for the eigengap that is independent of the mesh size h.

Proof of Theorem E.8. Suppose that there is no positive eigengap. Then $\lambda_h^1 \to \lambda^*$ when passing to a subsequence. Notice that

(E.8) $$\langle -\Delta_h \mathbf{u}^1, \mathbf{u}^1 \rangle_h + \langle \mathbb{V}\mathbf{u}^1, \mathbf{u}^1 \rangle_h + \beta\langle (\mathbf{u}^*)^2\mathbf{u}^1, \mathbf{u}^1 \rangle_h = \lambda_h^1,$$

which implies the boundedness of $|u_h^1|_{H_0^1(\Omega)} \leq \sqrt{\langle -\Delta_h \mathbf{u}^1, \mathbf{u}^1 \rangle_h}$, where $u_h^1 = \Pi\mathbf{u}^1$ and we used Lemma E.2 (iii). There exists $u^1 \in H_0^1(\Omega)$ such that u_h^1 converges to u^1, a.e., weakly in $H_0^1(\Omega)$, and strongly in $L^2(\Omega)$ and $L^4(\Omega)$, when passing to another subsequence.

Let us analyze the limiting behavior of each term in (E.8). It follows from the weak convergence $u_h^1 \to u^1$ in $H_0^1(\Omega)$ that

$$\liminf_{h\to 0}\langle -\Delta_h \mathbf{u}^1, \mathbf{u}^1 \rangle_h \geq \liminf_h \|u_h^1\|_{H_0^1(\Omega)}^2 \geq \|u^1\|_{H_0^1(\Omega)}^2.$$

By Fatou's lemma and Lemma E.2 (ii), one has

$$\liminf_{h\to 0}\langle \mathbb{V}\mathbf{u}^1, \mathbf{u}^1 \rangle_h = \liminf_{h\to 0} \int V|u_h^1|^2 \geq \int V|u^1|^2.$$

Let $\bar{u}_h^* = \bar{\Pi}\mathbf{u}^*$ and $\bar{u}_h^1 = \bar{\Pi}\mathbf{u}^1$ be the piecewise constant approximation of $\mathbf{u}^*$ and $\mathbf{u}^1$ in the sense of the one in Lemma E.5, then

$$\lim_{h\to 0} \|\bar{u}_h^* - u_h^*\|_{L^2(\Omega)} = 0 \quad \text{and} \quad \lim_{h\to 0} \|\bar{u}_h^1 - u_h^1\|_{L^2(\Omega)} = 0,$$

which implies $\bar{u}_h^* - u_h^* \to 0$ and $\bar{u}_h^1 - u_h^1 \to 0$ a.e. when passing to a subsequence. By Fatou's lemma, it holds that

$$\liminf_{h\to 0}\langle (\mathbf{u}^*)^2\mathbf{u}^1, \mathbf{u}^1 \rangle_h = \liminf_{h\to 0} \int |\bar{u}_h^*|^2|\bar{u}_h^1|^2 \geq \int |u^*|^2|u^1|^2.$$

Then taking limit for (E.8) as $h \to 0$, one obtains that

$$\lambda^* \geq \int |\nabla u^1|^2 + V|u^1|^2 + \beta|u^*|^2|u^1|^2 \geq \lambda^* \int |u^1|^2 = \lambda^*,$$

where the last equality follows from $\|u_h^1\|_{L^2(\Omega)} = 1 + o(1)$ by Lemma E.2 (i) and $u_h^1 \to u^1$ strongly in $L^2(\Omega)$. This implies that u^1 is also a ground state of the linearized problem at u^*. However, one has from Lemma E.2 (i) that

$$0 = \langle \mathbf{u}^*, \mathbf{u}^1 \rangle_h = 1 - \frac{1}{2}\|\mathbf{u}^* - \mathbf{u}_h^1\|_2^2 = 1 - \frac{1}{2}\|u_h^* - u_h^1\|_{L^2(\Omega)}^2 + O(h^2),$$

which implies that $0 = 1 - \frac{1}{2}\|u^* - u^1\|_{L^2(\Omega)}^2 = (u^*, u^1)_{L^2(\Omega)}$, leading to a contradiction. $\square$

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2008. With a foreword by Paul Van Dooren, DOI 10.1515/9781400830244. MR2364186

[2] R. Altmann, P. Henning, and D. Peterseim, *The J-method for the Gross-Pitaevskii eigenvalue problem*, Numer. Math. **148** (2021), no. 3, 575–610, DOI 10.1007/s00211-021-01216-5. MR4293963

[3] X. Antoine, W. Bao, and C. Besse, *Computational methods for the dynamics of the nonlinear Schrödinger/Gross-Pitaevskii equations*, Comput. Phys. Commun. **184** (2013), no. 12, 2621–2633, DOI 10.1016/j.cpc.2013.07.012. MR3128901

[4] I. Babuška and J. E. Osborn, *Estimates for the errors in eigenvalue and eigenvector approximation by Galerkin methods, with particular attention to the case of multiple eigenvalues*, SIAM J. Numer. Anal. **24** (1987), no. 6, 1249–1276, DOI 10.1137/0724082. MR917451

[5] I. Babuška and J. E. Osborn, *Finite element-Galerkin approximation of the eigenvalues and eigenvectors of selfadjoint problems*, Math. Comp. **52** (1989), no. 186, 275–297, DOI 10.2307/2008468. MR962210

[6] I. Babuška and J. Osborn, *Eigenvalue Problems*, Handbook of numerical analysis, Vol. II, Handb. Numer. Anal., II, North-Holland, Amsterdam, 1991, pp. 641–787. MR1115240

[7] W. Bao, I-L. Chern, and F. Y. Lim, *Efficient and spectrally accurate numerical methods for computing ground and first excited states in Bose-Einstein condensates*, J. Comput. Phys. **219** (2006), no. 2, 836–854, DOI 10.1016/j.jcp.2006.04.019. MR2274959

[8] W. Bao and Q. Du, *Computing the ground state solution of Bose-Einstein condensates by a normalized gradient flow*, SIAM J. Sci. Comput. **25** (2004), no. 5, 1674–1697, DOI 10.1137/S1064827503422956. MR2087331

[9] W. Bao, D. Jaksch, and P. A. Markowich, *Numerical solution of the Gross-Pitaevskii equation for Bose-Einstein condensation*, J. Comput. Phys. **187** (2003), no. 1, 318–342, DOI 10.1016/S0021-9991(03)00102-5. MR1977789

[10] W. Bao and W. Tang, *Ground-state solution of Bose-Einstein condensate by directly minimizing the energy functional*, J. Comput. Phys. **187** (2003), no. 1, 230–254, DOI 10.1016/S0021-9991(03)00097-4. MR1977785

[11] S. N. Bose, *Plancks Gesetz und Lichtquantenhypothese*, Zeits. Phys. **26** (1924), 178–181.

[12] E. Cancès, R. Chakir, L. He, and Y. Maday, *Two-grid methods for a class of nonlinear elliptic eigenvalue problems*, IMA J. Numer. Anal. **38** (2018), no. 2, 605–645, DOI 10.1093/imanum/drw053. MR3800034

[13] E. Cancès, R. Chakir, and Y. Maday, *Numerical analysis of nonlinear eigenvalue problems*, J. Sci. Comput. **45** (2010), no. 1-3, 90–117, DOI 10.1007/s10915-010-9358-1. MR2679792

[14] E. Cancès and C. Le Bris, *Can we outperform the DIIS approach for electronic structure calculations?*, Int. J. Quantum Chem. **79** (2000), no. 2, 82–90.

[15] E. Cancès and C. Le Bris, *On the convergence of SCF algorithms for the Hartree-Fock equations*, M2AN Math. Model. Numer. Anal. **34** (2000), no. 4, 749–774, DOI 10.1051/m2an:2000102. MR1784484

[16] Z. Chen, J. Lu, Y. Lu, and X. Zhang, *On the convergence of Sobolev gradient flow for the Gross-Pitaevskii eigenvalue problem*, SIAM J. Numer. Anal. **62** (2024), no. 2, 667–691, DOI 10.1137/23M1552553. MR4712406

[17] P. G. Ciarlet, M. H. Schultz, and R. S. Varga, *Numerical methods of high-order accuracy for nonlinear boundary value problems. III. Eigenvalue problems*, Numer. Math. **12** (1968), 120–133, DOI 10.1007/BF02173406. MR233517

[18] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, Classics in Applied Mathematics, vol. 40, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002. Reprint of the 1978 original [North-Holland, Amsterdam; MR0520174 (58 #25001)], DOI 10.1137/1.9780898719208. MR1930132

[19] L. J. Cross and X. Zhang, *On the monotonicity of Q^2 spectral element method for Laplacian on quasi-uniform rectangular meshes*, Commun. Comput. Phys. **35** (2024), no. 1, 160–180, DOI 10.4208/cicp.oa-2023-0206. MR4702092

[20] L. J. Cross and X. Zhang, *On the monotonicity of Q^3 spectral element method for Laplacian*, Ann. Appl. Math. **40** (2024), no. 2, 161–190. MR4746324

[21] F. Dalfovo, S. Giorgini, L. P Pitaevskii, and S. Stringari, *Theory of Bose-Einstein condensation in trapped gases*, Rev. Mod. Phys. **71** (1999), no. 3, 463.

[22] I. Danaila and P. Kazemi, *A new Sobolev gradient method for direct minimization of the Gross-Pitaevskii energy with rotation*, SIAM J. Sci. Comput. **32** (2010), no. 5, 2447–2467, DOI 10.1137/100782115. MR2684722

[23] I. Danaila and B. Protas, *Computation of ground states of the Gross-Pitaevskii functional via Riemannian optimization*, SIAM J. Sci. Comput. **39** (2017), no. 6, B1102–B1129, DOI 10.1137/17M1121974. MR3731036

[24] M. Defranceschi and C. Le Bris (eds.), *Mathematical models and methods for ab initio quantum chemistry*, Lecture Notes in Chemistry, vol. 74, Springer-Verlag, Berlin, 2000, DOI 10.1007/978-3-642-57237-1. MR1857459

[25] Y. Du, *Order Structure and Topological Methods in Nonlinear Partial Differential Equations. Vol. 1*, Series in Partial Differential Equations and Applications, vol. 2, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2006. Maximum principles and applications, DOI 10.1142/9789812774446. MR2205529

[26] G. Dusson and Y. Maday, *An overview of a posteriori error estimation and post-processing methods for nonlinear eigenvalue problems*, J. Comput. Phys. **491** (2023), Paper No. 112352, 13, DOI 10.1016/j.jcp.2023.112352. MR4622467

[27] A. Einstein, *Quantentheorie des einatomigen idealen Gases (Zweite Abhandlung)*, SB Preuss. Akad. Wiss. phys-math. Klasse **13** (1925), 3–14.

[28] D. Gallistl, M. Hauck, Y. Liang, and D. Peterseim, *Mixed finite elements for the Gross-Pitaevskii eigenvalue problem: a priori error analysis and guaranteed lower energy bound*, Preprint, `arXiv:2402.06311`, 2024.

[29] P. Gravejat, *Decay for travelling waves in the Gross-Pitaevskii equation* (English, with English and French summaries), Ann. Inst. H. Poincaré C Anal. Non Linéaire **21** (2004), no. 5, 591–637, DOI 10.1016/j.anihpc.2003.09.001. MR2086751

[30] M. Hauck, Y. Liang, and D. Peterseim, *Positivity preserving finite element method for the gross-pitaevskii ground state: discrete uniqueness and global convergence*, Preprint, `arXiv:2405.17090`, 2024.

[31] P. Heid, B. Stamm, and T. P. Wihler, *Gradient flow finite element discretizations with energy-based adaptivity for the Gross-Pitaevskii equation*, J. Comput. Phys. **436** (2021), Paper No. 110165, 15, DOI 10.1016/j.jcp.2021.110165. MR4234216

[32] P. Henning, *The dependency of spectral gaps on the convergence of the inverse iteration for a nonlinear eigenvector problem*, Math. Models Methods Appl. Sci. **33** (2023), no. 7, 1517–1544, DOI 10.1142/S0218202523500343. MR4595654

[33] P. Henning and D. Peterseim, *Sobolev gradient flow for the Gross-Pitaevskii eigenvalue problem: global convergence and computational efficiency*, SIAM J. Numer. Anal. **58** (2020), no. 3, 1744–1772, DOI 10.1137/18M1230463. MR4108207

[34] J. Hu and X. Zhang, *Positivity-preserving and energy-dissipative finite difference schemes for the Fokker-Planck and Keller-Segel equations*, IMA J. Numer. Anal. **43** (2023), no. 3, 1450–1484, DOI 10.1093/imanum/drac014. MR4597190

[35] E. Jarlebring, S. Kvaal, and W. Michiels, *An inverse iteration method for eigenvalue problems with eigenvector nonlinearities*, SIAM J. Sci. Comput. **36** (2014), no. 4, A1978–A2001, DOI 10.1137/130910014. MR3249374

[36] P. Kazemi and M. Eckart, *Minimizing the Gross-Pitaevskii energy functional with the Sobolev gradient—analytical and numerical results*, Int. J. Comput. Methods **7** (2010), no. 3, 453–475, DOI 10.1142/S0219876210002301. MR2736464

[37] A. V. Knyazev and J. E. Osborn, *New a priori FEM error estimates for eigenvalues*, SIAM J. Numer. Anal. **43** (2006), no. 6, 2647–2667, DOI 10.1137/040613044. MR2206452

[38] A. de Laire and S. López-Martínez, *Existence and decay of traveling waves for the nonlocal Gross-Pitaevskii equation*, Comm. Partial Differential Equations **47** (2022), no. 9, 1732–1794, DOI 10.1080/03605302.2022.2070853. MR4472922

[39] H. Li, D. Appelö, and X. Zhang, *Accuracy of spectral element method for wave, parabolic, and Schrödinger equations*, SIAM J. Numer. Anal. **60** (2022), no. 1, 339–363, DOI 10.1137/21M1401760. MR4378595

[40] H. Li, S. Xie, and X. Zhang, *A high order accurate bound-preserving compact finite difference scheme for scalar convection diffusion equations*, SIAM J. Numer. Anal. **56** (2018), no. 6, 3308–3345, DOI 10.1137/18M1208551. MR3880256

[41] H. Li and X. Zhang, *On the monotonicity and discrete maximum principle of the finite difference implementation of C^0-Q^2 finite element method*, Numer. Math. **145** (2020), no. 2, 437–472, DOI 10.1007/s00211-020-01110-6. MR4107225

[42] H. Li and X. Zhang, *Superconvergence of high order finite difference schemes based on variational formulation for elliptic equations*, J. Sci. Comput. **82** (2020), no. 2, Art. 36, 39, DOI 10.1007/s10915-020-01144-w. MR4058439

[43] H. Li and X. Zhang, *A high order accurate bound-preserving compact finite difference scheme for two-dimensional incompressible flow*, Commun. Appl. Math. Comput. **6** (2024), no. 1, 113–141, DOI 10.1007/s42967-022-00227-9. MR4710829

[44] E. H. Lieb, R. Seiringer, and J. Yngvason, *Bosons in a Trap: A rigorous derivation of the Gross-Pitaevskii energy functional*, The Stability of Matter: From Atoms to Stars: Selecta of Elliott H. Lieb, 2001, pp. 685–697.

[45] X. Liu, J. Shen, and X. Zhang, *A simple GPU implementation of spectral-element methods for solving 3D Poisson type equations on rectangular domains and its applications*, to appear in Commun. Computat. Phys., DOI 10.4208/cicp.OA-2024-0072.

[46] L. Pitaevskii and S. Stringari, *Bose-Einstein Condensation*, International Series of Monographs on Physics, vol. 116, The Clarendon Press, Oxford University Press, Oxford, 2003. MR2012737

[47] R. J. Plemmons, *M-matrix characterizations. I. Nonsingular M-matrices*, Linear Algebra Appl. **18** (1977), no. 2, 175–188, DOI 10.1016/0024-3795(77)90073-8. MR444681

[48] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. III*, Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1979. Scattering theory. MR529429

[49] J. Shen and X. Zhang, *Discrete maximum principle of a high order finite difference scheme for a generalized Allen-Cahn equation*, Commun. Math. Sci. **20** (2022), no. 5, 1409–1436. MR4436138

[50] P. Upadhyaya, E. Jarlebring, and E. H. Rubensson, *A density matrix approach to the convergence of the self-consistent field iteration*, Numer. Algebra Control Optim. **11** (2021), no. 1, 99–115, DOI 10.3934/naco.2020018. MR4236696

[51] R. S Varga, *Matrix Iterative Analysis*, Vol. 27, Springer Science & Business Media, 1999.

[52] J. Xu and L. Zikatanov, *A monotone finite element scheme for convection-diffusion equations*, Math. Comp. **68** (1999), no. 228, 1429–1446, DOI 10.1090/S0025-5718-99-01148-5. MR1654022

[53] Z. Zhang, *Exponential convergence of Sobolev gradient descent for a class of nonlinear eigenproblems*, Commun. Math. Sci. **20** (2022), no. 2, 377–403, DOI 10.4310/CMS.2022.v20.n2.a4. MR4374290

Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139
    Email address: ziang@mit.edu

Departments of Mathematics, Physics, and Chemistry, Duke University, Box 90320, Durham, North Carolina 27708
    Email address: jianfeng@math.duke.edu

School of Mathematics, University of Minnesota, 206 Church Street SE, Minneapolis, Minnesota 55455
    Email address: yulonglu@umn.edu

Department of Mathematics, Purdue University, 150 N. University Street, West Lafayette, Indiana 47907
    Email address: zhan1966@purdue.edu