COMMUN. MATH. SCI.
Vol. 20, No. 2, pp. 377–403

EXPONENTIAL CONVERGENCE OF SOBOLEV GRADIENT DESCENT FOR A CLASS OF NONLINEAR EIGENPROBLEMS*

ZIYUN ZHANG†

Abstract. We propose to use the Łojasiewicz inequality as a general tool for analyzing the convergence rate of gradient descent on a Hilbert manifold, without resorting to the continuous gradient flow. Using this tool, we show that a Sobolev gradient descent method with adaptive inner product converges exponentially fast to the ground state for the Gross-Pitaevskii eigenproblem. This method can be extended to a class of general high-degree optimizations or nonlinear eigenproblems under certain conditions. We demonstrate this generalization using several examples, in particular a nonlinear Schrödinger eigenproblem with an extra high-order interaction term. Numerical experiments are presented for these problems.

Keywords. Sobolev gradient descent, Gross-Pitaevskii eigenproblem, Łojasiewicz inequality, Schrödinger equation, nonlinear eigenproblems.

AMS subject classifications. 35P30; 47J10; 65K10; 65N25; 81Q05.

1. Introduction

The Gross-Pitaevskii eigenproblem, a well-known example of the nonlinear Schrödinger eigenproblem, seeks $\lambda\in\mathbb{R}$ and $v\in H_0^1(\Omega)$ that satisfy the following equation

$$-\Delta v+Vv+\beta|v|^2v=\lambda v \quad \text{on } \Omega\subset\mathbb{R}^d, \tag{1.1}$$

where Ω is a bounded region in $\mathbb{R}^d$, $V(x)\geq 0$ is an external trapping potential, and $\beta\geq 0$ is a parameter describing the repulsive interaction between particles. In physics, this describes the Bose-Einstein condensate when the temperature is close to absolute zero. The eigenstate v corresponding to the smallest λ describes the ground state of this system. It has long been studied both in experiments [2] and in numerical analysis [8, 16, 22, 26].

To find the ground state v is equivalent to solving the following minimization problem:

$$\min_{\|u\|_{L^2}=1,\, u\in H_0^1(\Omega)} E(u):=\int_\Omega\left(|\nabla u|^2+V|u|^2+\frac{\beta}{2}|u|^4\right)\mathrm{d}x. \tag{1.2}$$

The constraint set $\{u\in H_0^1(\Omega):\|u\|_{L^2}=1\}$ is the unit sphere in $H_0^1(\Omega)$. It can be seen as an infinite dimensional Hilbert manifold. Such a manifold (with additional $L^\infty(\Omega)$ constraints) will be denoted as $\mathcal{M}$ in subsequent sections. Thus many manifold optimization methods on the Riemannian manifold are readily applicable to this problem, with diverse techniques and rich theories.

In this paper, we focus on a special manifold gradient descent method named the *Sobolev projected gradient descent (Sobolev PGD)*, first proposed in [23]. This method has the following iteration formula:

$$u_{n+1}=R\left((1-\tau_n)u_n+\tau_n\cdot\frac{(u_n,u_n)_{L^2}}{(\mathcal{G}_{u_n}u_n,\mathcal{G}_{u_n}u_n)_{a_{u_n}}}\mathcal{G}_{u_n}u_n\right), \tag{1.3}$$

*Received: September 23, 2020; Accepted (in revised form): July 10, 2021. Communicated by Weizhu Bao.

†Applied and Computational Mathematics, Caltech, Pasadena, CA 91125, USA (zyzhang@caltech.edu).

where R is the retraction back onto the manifold, τ_n is the n-th step size, $(\cdot,\cdot)_{a_{u_n}}$ is an adaptive inner product in the tangent space of $\mathcal{M}$, and $\mathcal{G}_{u_n}$ is the Green's operator associated with $(\cdot,\cdot)_{a_{u_n}}$. Their definitions are in Section 3. The main result of this paper is as follows.

THEOREM 1.1 (Main result, informal). *If initialized with a positive initial guess u_0, the a_u-Sobolev gradient descent which is given by (1.3) converges to the ground state of the eigenproblem (1.1) exponentially fast.*

The idea of using a discretized normalized gradient flow (DNGF) to solve Problem (1.2) can be traced back to [6]. Following this seminal work there have been a number of variants, see e.g. [12,13,17] and the review paper [4]. The viewpoint of (Riemannian) manifold optimization has also been explicitly adopted in [13]. Based on those methods with fixed inner products, the adaptive version of a_u-Sobolev gradient descent has recently been proposed in [23]. Despite its popularity, quantitative convergence analysis of the DNGF family has been quite lacking. The convergence rate has been either unavailable, or only proved for the gradient flow [23]. Another popular choice is the self consistent field iteration (SCF), see e.g. [10]. Rigorous global convergence rate is however difficult to establish. There are also second-order methods like the Riemannian Newton method, but they require second-order information which can be expensive to obtain.

We highlight the main differences between the current paper and [23]. The authors of [23] first propose the Sobolev gradient descent method (1.3). They establish the exponential convergence rate of the *time-continuous* gradient flow. But the important question of whether the *time-discrete* gradient descent also achieves optimal exponential convergence rate remains open. Our main contribution is to give a confirmatory answer to this question. We do this by introducing the *Łojasiewicz inequality* tool, which is a general analytical tool that is applicable to a wide class of problems.

Specifically, in Section 2, using the Łojasiewicz inequality tool, we reveal that the key to exponential convergence is the quadratic nature of the objective energy functional. In other words, regarded as a polynomial, the objective functional should behave like a degree-2 polynomial under the given manifold metric. The Łojasiewicz inequality has been widely used in the optimization community, see e.g. [19, 27]. Yet it has scarcely been applied to the problems of interest in this paper.

Although the degree of polynomial of the objective function in Problem (1.2) is formally higher than quadratic, Method (1.3) changes the situation by using an adaptive inner product $a_u(\cdot,\cdot)$ instead of a fixed inner product. As a comparison, using a fixed inner product, the Łojasiewicz exponent (the θ in Theorem 2.1) calculated in [28] is $1/4$; while in this paper, using an adaptive inner product, we have $\theta = 1/2$. The latter is more desirable according to Theorem 2.1. Thus, in Section 3, using the Łojasiewicz inequality tool, we are able to prove the exponential convergence rate of discrete time gradient descent directly.

The Łojasiewicz inequality tool also makes the Sobolev gradient descent easily applicable to general optimization of high-degree objective or eigenvalue problems other than the Gross-Pitaevskii eigenvalue problem. Its interesting property of making a high-degree polynomial behave like quadratic is not specific to a certain problem, but is general. Examples include the biharmonic Schrödinger, the nonlinear Schrödinger with a different order or extra interaction terms, and potentially some general manifold optimization problems.

In addition to the necessary regularity conditions, the only essential requirement is

that the global ground state of the nonlinear problem is also the unique ground state of its *linearized* version, what we call the "double ground state" property.[1] For Problem (1.1), this property will be rigorously proved in Section 3.2. For many other problems, it is either provable, or a reasonable assumption according to numerical evidence. We summarize this result as the following:

PROPOSITION 1.1 (Generalization of main result, informal). *If the objective problem satisfies the "double ground state" property and necessary regularity conditions, then with a proper initialization u_0, the a_u-Sobolev gradient descent converges to a minimizer of this problem exponentially fast.*

Specifically, an example of nonlinear Schrödinger eigenproblem from [5] will be rigorously discussed in Section 5. This example has an extra high-order interaction term $-\delta\Delta(|v|^2)v$ where $\delta \geq 0$. Classical methods that work for (1.1) could become inefficient or unstable for this problem. A density function reformulation $\rho := |u|^2$ was proposed in [7], but it has to treat the lack of continuity of $\nabla\sqrt{\rho}$ near 0^+ with extra regularization. Therefore the adaptive Sobolev gradient descent is advantageous for its simplicity and fast convergence.

We remark that if the domain is convex, an alternative approach to derive local linear convergence rate[2] of gradient descent methods is to use *strong convexity* (SC). This is especially popular in the finite dimensional data science problems [11]. Attempts have also been made to extend it to nonconvex settings like manifolds. Some works in this direction can be found in [1,9]. We emphasize that our approach using the Łojasiewicz inequality has its advantages over SC, namely it applies to degenerate critical points where SC could fail, and it allows more freedom in the choice of iterative algorithms and convergence measures. A more detailed comparison of these two approaches would be of interest in future research.

The rest of the paper is organized as follows. In Section 2, we introduce the Łojasiewicz inequality tool with mixed norms on the Hilbert manifold as an abstract convergence theorem. In Section 3, we establish the main result on the exponential convergence of the a_u-Sobolev gradient descent method applied to the Gross-Pitaevskii eigenproblem (1.1). Section 4 is devoted to the analysis of spatial discretization. In Section 5, we introduce several extensions of the Sobolev gradient descent to other nonlinear eigenproblems. Some numerical results are presented in Section 6. Finally, we make some concluding remarks in Section 7.

2. Abstract convergence theorem using the Łojasiewicz inequality

In this section, we introduce the Łojasiewicz inequality tool as an abstract convergence theorem. We show that one can deduce the convergence of an iteration algorithm from a triplet of conditions (L), (D) and (S). Furthermore, whether the convergence rate is exponential (linear) or polynomial (sublinear) is determined by the exponent in the (L) inequality.

THEOREM 2.1. *Assume that the domain $\mathcal{M}$ is a Hilbert manifold. Let $\|\cdot\|_X$ be a norm on $\mathcal{TM}$, the tangent bundle of $\mathcal{M}$, and $\|\cdot\|_Y$ be a norm in the ambient space of $\mathcal{M}$ which is complete. Here $\|\cdot\|_X$ and $\|\cdot\|_Y$ can be either same or different. Let*

[1] This property is nontrivial. Although an eigenstate of the nonlinear problem is always an eigenstate of the linearized problem, it is not always the *lowest energy* eigenstate (i.e., ground state) of the linearized problem.

[2] Both *exponential* and *linear convergence* refer to the case where $err_k \leq c^k \cdot err_0$ for some $0 < c < 1$. In this paper we use both terms interchangeably. The term *linear convergence* is more popular in the optimization community.

$\{u_n\}_{n=0}^{\infty} \subset \mathcal{M}$ be a sequence generated by some iterative algorithm. Assume that $E(u)$ is differentiable on $\mathcal{M}$ and let grad $E(u)$ be the manifold gradient of $E(u)$. If $E(u)$ and $\{u_n\}_{n=0}^{\infty}$ satisfy the following conditions for all $n \in \mathbb{Z}_+$:

- (Łojasiewicz Gradient Inequality.) *There exists u^* that is a cluster point of $\{u_n\}$, and there exists $0 < C_L < +\infty$, $0 < \theta \le \frac{1}{2}$, such that for large enough n,*

$$|E(u_n) - E(u^*)|^{1-\theta} \le C_L \|grad\ E(u_n)\|_X; \tag{L}$$

- (Descent Inequality.) *There exists $C_D > 0$ such that for large enough n,*

$$E(u_n) - E(u_{n+1}) \ge C_D \|grad\ E(u_n)\|_X \|u_{n+1} - u_n\|_Y; \tag{D}$$

- (Step-size Condition.) *There exists $C_S > 0$ such that for large enough n,*

$$\|u_{n+1} - u_n\|_Y \ge C_S \|grad\ E(u_n)\|_X. \tag{S}$$

Then u^* is the unique limit point of $\{u_n\}_{n=0}^{\infty}$ w.r.t. $\|\cdot\|_Y$. Moreover, $\{u_n\}_{n=0}^{\infty}$ converge to u^* with the following asymptotic convergence rate:

$$\|u_n - u^*\|_Y \lesssim \begin{cases} e^{-cn}, & \text{if } \theta = \frac{1}{2}, \\ n^{-\frac{\theta}{1-2\theta}}, & \text{if } \theta \in (0, \frac{1}{2}), \end{cases}$$

where $c := \log(1 - \frac{C_D C_S}{2C_L^2})$.

Proof. $\{E(u_n)\}$ is monotonically decreasing from Condition (D). Since u^* is a cluster point of $\{u_n\}$, $E(u_n) \ge E(u^*)$ for any n. We also have $\lim_{n\to\infty} E(u_n) = E(u^*)$ by continuity of $E(\cdot)$. Without loss of generality, assume that $E(u^*) = 0$. By Conditions (D) and (L), we have

$$\begin{aligned} \|u_{n+1} - u_n\|_Y &\le \frac{E(u_n) - E(u_{n+1})}{C_D \|\text{grad } E(u_n)\|_X} \le \frac{C_L}{C_D} (E(u_n) - E(u_{n+1})) E(u_n)^{\theta - 1} \\ &\le \frac{C_L}{C_D} \int_{E(u_{n+1})}^{E(u_n)} y^{\theta-1}\, \mathrm{d}y = \frac{C_L}{\theta C_D} (E(u_n)^{\theta} - E(u_{n+1})^{\theta}). \end{aligned}$$

Using a bootstrapping argument, we have that for any $m > n$,

$$\|u_n - u_m\|_Y \le \frac{C_L}{\theta C_D} (E(u_n)^{\theta} - E(u_m)^{\theta}) \le \frac{C_L}{\theta C_D} E(u_n)^{\theta}. \tag{2.1}$$

Since $E(u_n)$ is convergent, we deduce that u_n is convergent, and the limit point is u^*.

To estimate the convergence rate, let $r_n := \sum_{k=n}^{\infty} \|u_{k+1} - u_k\|_Y$, then $\|u_n - u^*\|_Y \le r_n$. It suffices to estimate the convergence rate of r_n. By Conditions (L) and (S), for large enough n,

$$|E(u_n) - E(u^*)|^{1-\theta} \le C_L \|\text{grad } E(u_n)\|_X \le \frac{C_L}{C_S} \|u_{n+1} - u_n\|_Y.$$

Since we have made the assumption that $E(u^*) = 0$, we obtain

$$E(u_n) \le \left(\frac{C_L}{C_S} \|u_{n+1} - u_n\|_Y \right)^{\frac{1}{1-\theta}}. \tag{2.2}$$

Thus, we have

$$r_n = \sum_{k=n}^{\infty} \|u_{k+1}-u_k\|_Y \le \sum_{k=n}^{\infty} \frac{C_L}{\theta C_D}(E(u_k)^\theta - E(u_{k+1})^\theta) = \frac{C_L}{\theta C_D} E(u_n)^\theta$$
$$\le \frac{C_L}{\theta C_D}\left(\frac{C_L}{C_S}\|u_{n+1}-u_n\|_Y\right)^{\frac{\theta}{1-\theta}} = \frac{C_L}{\theta C_D}\left(\frac{C_L}{C_S}(r_n - r_{n+1})\right)^{\frac{\theta}{1-\theta}},$$

where the first inequality is due to (2.1) and the second inequality is due to (2.2). This gives

$$r_{n+1} \le r_n - C r_n^{\frac{1-\theta}{\theta}}, \quad C := C_L^{-\frac{1}{\theta}} (\theta C_D)^{\frac{1-\theta}{\theta}} C_S.$$

Note that here $0 < C < 1$, otherwise the sequence would have converged in finite steps.

If $\theta \in (0, \frac{1}{2})$, let $s_n := s_0 n^{-\gamma}$, $\gamma = \frac{\theta}{1-2\theta}$, and $s_0 \ge \max\{r_0, (C/\gamma)^{-\gamma}\}$. Then

$$s_{n+1} = s_n\left(1+\frac{1}{n}\right)^{-\gamma} \ge s_n\left(1-\frac{1}{n}\cdot\gamma\right) = s_n\left(1-\gamma s_0^{-1/\gamma} s_n^{1/\gamma}\right) \ge s_n - C s_n^{\frac{\gamma+1}{\gamma}} = s_n - C s_n^{\frac{1-\theta}{\theta}}.$$

Combining $s_0 \ge r_0$, $r_{n+1} \le r_n - C r_n^{\frac{1-\theta}{\theta}}$, and $s_{n+1} \ge s_n - C s_n^{\frac{1-\theta}{\theta}}$, by induction,

$$r_n \le s_n = s_0 n^{-\frac{\theta}{1-2\theta}} \quad \forall n,$$

which is polynomial (or sub-linear) convergence.

If $\theta = \frac{1}{2}$, then $r_{n+1} \le (1-C) r_n$, and

$$r_n \le r_0 e^{cn}, \quad c := \ln(1-C),$$

which is exponential (or linear) convergence. ∎

The above result can be seen as a generalization of Theorem 2.3 in [27] to the Hilbert space/manifold. Another work in this direction is [19]. What is new in our version is that one has the freedom to choose mixed norms ($\|\cdot\|_X$ and $\|\cdot\|_Y$), as long as the conditions (L), (D) and (S) can be satisfied under these norms. One example is the $\|\cdot\|_{a_u}$ in this paper, which varies with u.

The advantage of the Łojasiewicz inequality approach is that instead of dealing with the time discretization of the gradient flow, it gives the convergence of the gradient descent directly. The triplet of conditions (L), (D) and (S) in Theorem 2.1 all have clear and intuitive meanings. In fact, it is easier to deduce the convergence property of the gradient flow from that of the gradient descent, since we only need to take the limit $\tau \to 0^+$; while the reverse direction from gradient flow to gradient descent can be more difficult.

An important observation is that the exponent θ in Łojasiewicz gradient inequality indicates the *degree of polynomial* of the objective function. For example, consider $x \in \mathbb{R}$, let $f(x) = x^k$ for a positive integer k, then Łojasiewicz gradient inequality holds with $\theta = 1/k$. From this viewpoint, exponential convergence is closely related to certain quadratic-like behavior of the objective functional. It is thus unusual for a quartic-quadratic functional $E(\cdot)$ (i.e. a functional which is the sum of nonnegative quartic and quadratic terms) to have exponential convergence rate. What the Sobolev gradient does is to force the quartic term to behave like quadratic. This is the idea behind the proof of Theorem 3.2.

3. Exponential convergence of Sobolev gradient descent

In this section, we establish the convergence rate of the a_u-Sobolev gradient descent for Problems (1.1) and (1.2). In Section 3.1, we introduce the setting of manifold optimization and derive the a_u-Sobolev gradient descent method. In Section 3.2, using the Łojasiewicz inequality tool from the previous section, we prove the exponential convergence rate by checking conditions (L), (D) and (S) for this specific method.

3.1. Manifold setting and derivation of a_u-Sobolev gradient descent.

The following assumptions on Ω, V and β will be required throughout this section.

ASSUMPTION 3.1. *Let Ω, V and β be chosen such that the following assumptions hold:*

- *Ω is a bounded domain in $\mathbb{R}^d$, $d=1,2$, or 3, and Ω is either convex Lipschitz or has a smooth boundary;*
- *$V\geq 0$ and $V\in L^\infty(\Omega)$, V is a trapping potential, and $\beta\geq 0$.*

REMARK 3.1. V is chosen as a trapping potential so that the eigenstates of interest are localized. It is then natural to impose zero Dirichlet boundary conditions on $\partial\Omega$. Examples of a trapping potential include the well model in the classical Anderson localization where $\lim_{|x|\to\infty} V(x)=+\infty$, and the fully disordered model with high contrast and small interaction length.

Define the infinite dimensional Hilbert manifold $\mathcal{M}$ as

$$\mathcal{M}:=\{u\in H_0^1(\Omega):\|u\|_{L^2(\Omega)}=1,\|u\|_{L^\infty(\Omega)}\leq M_0 \text{ for some global constant } M_0\}.$$

Then $\mathcal{M}$ is a submanifold in $H_0^1(\Omega)\cap L^\infty(\Omega)$. Note that although the original problem (1.1) allows $v(x)\in\mathbb{C}$, we restrict our search to $u(x)\in\mathbb{R}$, as we will see that the existence of a real and positive ground state is ensured by Theorem 3.1. We also remark that $\|u\|_{L^\infty(\Omega)}\leq M_0$ is not directly guaranteed by the iterative algorithm, but is rather left as an assumption. It is a plausible assumption because we will see that the ground state v is in $L^\infty(\Omega)$ by Hölder continuity in Theorem 3.1. For simplicity we drop Ω in norm and inner product notations when there is no confusion. The *tangent space* of $\mathcal{M}$ at point $u\in\mathcal{M}$ is defined as

$$\mathcal{T}_u\mathcal{M}=\{\xi\in H_0^1(\Omega)\cap L^\infty(\Omega):(\xi,u)_{L^2}=0\}. \tag{3.1}$$

We need an inner product in the tangent space, denoted as $(\cdot,\cdot)_X$. On the finite dimensional Riemannian manifold, this is dubbed the *Riemannian metric*. It can be easily generalized to the infinite dimensional Hilbert manifold.

For $u\neq 0$, the *retraction* of u onto $\mathcal{M}$ is given by

$$R(u)=u/\|u\|_{L^2}.$$

Note that the retraction operation itself is independent of the choice of the inner product $(\cdot,\cdot)_X$, but its approximation property is not. When the inner product $(\cdot,\cdot)_X$ is introduced, it is usually required that the retraction is at least first-order, i.e., $R(z+\xi)=z+o(\|\xi\|_X)$ for $z\in\mathcal{M}$ and $\xi\in\mathcal{T}_u\mathcal{M}$.

Given an inner product $(\cdot,\cdot)_X$, let $\mathcal{G}$ be its associated Green's operator, i.e.,

$$(z,\mathcal{G}w)_X=(z,w)_{L^2},\qquad \forall z,w\in X.$$

For an arbitrary element ξ in the ambient space, the *projection onto the tangent space* at point $u\in\mathcal{M}$ is given by

$$P_{\mathcal{T}_u\mathcal{M}}(\xi)=\xi-\frac{(\xi,u)_{L^2}}{(\mathcal{G}u,\mathcal{G}u)_X}\mathcal{G}u.$$

Given a differentiable function $E(u)$ defined on $\mathcal{M}$, the *Sobolev gradient* of $E(u)$ with respect to the inner product $(\cdot,\cdot)_X$ is the unique element $\nabla_X E(u) \in X$ such that

$$(\nabla_X E(u), w)_X = (\nabla E(u), w)_{L^2}, \qquad \forall w \in X.$$

The *manifold gradient* of $E(u)$ on $\mathcal{M}$, denoted as grad$E(u)$, is the projection of the Sobolev gradient onto the tangent space with respect to the inner product $(\cdot,\cdot)_X$. Thus we have

$$\text{grad } E(u) = P_{\mathcal{T}_u\mathcal{M}}(\nabla_X E(u)) = \nabla_X E(u) - \frac{(\nabla_X E(u), u)_{L^2}}{(\mathcal{G}u, \mathcal{G}u)_X}\mathcal{G}u.$$

It can be inferred from the above expression that grad $E(u) = 0$ implies $\nabla E(u) = \lambda u$ for some scalar λ. If $E(u)$ is as in (1.2), then u is an eigenstate of (1.1). This fact is independent of the choice of inner product $(\cdot,\cdot)_X$.

The choice of the inner product in the tangent space plays an important role in the analysis of manifold optimization algorithms as different inner products give different forms of gradient flow and gradient descent algorithms. Popular choices include L^2, H^1, and the a_0 inner product defined as follows:

$$(z,w)_{a_0} := \int_\Omega \nabla z \nabla w + V z w, \qquad \forall z, w \in \mathcal{T}_u\mathcal{M}, \quad u \in \mathcal{M}.$$

All the above inner products are fixed everywhere on the manifold. Things become interesting when the inner product becomes adapted to u. Specifically, we are interested in the following inner product

$$(z,w)_{a_u} := \int_\Omega \nabla z \nabla w + V z w + \beta |u|^2 z w, \qquad \forall z, w \in \mathcal{T}_u\mathcal{M}, \quad u \in \mathcal{M}, \tag{3.2}$$

and we define

$$\mathcal{A}_u := -\Delta + V + \beta |u|^2, \tag{3.3}$$

such that $(\mathcal{A}_u z, w)_{L^2} = (z,w)_{a_u}$ for any z, w. This new inner product $(\cdot,\cdot)_{a_u}$ can be seen as the linearization of the Gross-Pitaevskii energy functional. A desirable property of this inner product is that the Sobolev gradient of $E(u)$ is u itself, i.e.,

$$\nabla_{a_u} E(u) = u. \tag{3.4}$$

This inner product has the associated Green's operator $\mathcal{G}_u$ whose properties have been explored in [23].

LEMMA 3.1. *Under the adaptive inner product $(\cdot,\cdot)_{a_u}$, the retraction R is second-order.*

Proof. For $u \in \mathcal{M}$ and for any $\xi \in \mathcal{T}_u\mathcal{M}$,

$$\frac{\|R(u+\xi) - (u+\xi)\|_{a_u}}{\|u+\xi\|_{a_u}} = \frac{\|(1 - 1/\|u+\xi\|_{L^2})(u+\xi)\|_{a_u}}{\|u+\xi\|_{a_u}} = \left|1 - \frac{1}{\|u+\xi\|_{L^2}}\right|.$$

Note that ξ is a tangent vector of the manifold at u. By (3.1), $\|u+\xi\|_{L^2}^2 = \|u\|_{L^2}^2 + \|\xi\|_{L^2}^2 + 2(\xi,u)_{L^2} = 1 + \|\xi\|_{L^2}^2$. Thus we have

$$\frac{\|R(u+\xi) - (u+\xi)\|_{a_u}}{\|u+\xi\|_{a_u}} = \left|1 - (1 + \|\xi\|_{L^2}^2)^{-1/2}\right| = \frac{1}{2}\|\xi\|_{L^2}^2 + \mathcal{O}(\|\xi\|_{L^2}^4).$$

By the Poincaré inequality, when $V \geq 0$ and $\beta \geq 0$,

$$\|\xi\|_{L^2}^2 \leq C_P \|\nabla \xi\|_{L^2}^2 \leq C_P \|\xi\|_{a_u}^2$$

for some domain constant $C_P > 0$. Thus we have

$$\|R(u+\xi) - (u+\xi)\|_{a_u} = \mathcal{O}(\|\xi\|_{a_u}^2),$$

where the constant in $\mathcal{O}(\cdot)$ is independent of ξ. ∎

Using the inner product $(\cdot,\cdot)_{a_u}$, the manifold gradient becomes

$$\text{grad } E(u) = u - \frac{(u,u)_{L^2}}{(\mathcal{G}_u u, \mathcal{G}_u u)_{a_u}} \mathcal{G}_u u. \tag{3.5}$$

We now have the Sobolev projected gradient descent (Sobolev PGD) as in (1.3):

$$\begin{aligned} u_{n+1} &= R(u_n - \tau_n \cdot \text{grad } E(u_n)) \\ &= R\left((1-\tau_n) u_n + \tau_n \cdot \frac{(u_n,u_n)_{L^2}}{(\mathcal{G}_{u_n} u_n, \mathcal{G}_{u_n} u_n)_{a_{u_n}}} \mathcal{G}_{u_n} u_n \right). \end{aligned} \tag{3.6}$$

3.2. Asymptotic convergence and exponential rate. Throughout the rest of the paper, let v always denote the global minimizer of $E(u)$, i.e. the ground state of the nonlinear eigenproblem. Let λ always denote its corresponding eigenvalue. We have the following basic observations about the ground state v.

THEOREM 3.1. *There is a ground state v that satisfies $v(x) > 0$ everywhere on Ω. It is the only strictly positive eigenstate of (1.1) up to scaling. Moreover, it is both the* unique *ground state of the nonlinear eigenproblem (1.1) and the* unique *ground state of the linearized operator $\mathcal{A}_v$ up to the sign. Moreover, v has Hölder regularity $v \in C^{0,\alpha}(\bar{\Omega})$ for some $0 < \alpha < 1$.*

Proof. This theorem is a consequence of Lemma 2 in [8] and Lemmas 5.3 and 5.4 in [23]. We only outline the main idea of the proof here to make this paper self-contained.

The idea is that the existence of at least one global minimizer v is ensured by the convexity of $E(u)$. The Hölder continuity of v is ensured by elliptic regularity, see e.g. [21, Theorem 8.24]. This v can always be chosen to be nonnegative because $E(u) = E(|u|)$. This nonnegativity can be made into positivity by applying the Harnack inequality to $(A_v - \lambda)$, see e.g. [21, Corollary 8.21]. Thus, there exists a ground state of the nonlinear problem that is positive. The same argument shows that the ground state eigenfunction of the linearized operator $\mathcal{A}_v$ is also positive and is unique. Since v is an eigenfunction of $\mathcal{A}_v$ and is positive, it is exactly that ground state. Thus we have the "double ground state" property. Finally, the uniqueness of any positive eigenstate of the original nonlinear eigenproblem can be established by contradiction. This can be done either by the Picone identity as in [23], or by showing that as long as some u itself is the ground state of the linearized operator $\mathcal{A}_u$, it must be the ground state of the original problem. ∎

It turns out in subsequent results that v being the "double" ground state in Theorem 3.1 is essential to the exponential convergence rate.

LEMMA 3.2. *If the initial point u_0 of the Sobolev PGD satisfies $u_0 > 0$ everywhere on Ω, then $\{u_n\}_{n=0}^{\infty}$ generated by the Sobolev PGD with step size $\tau_{min} \leq \tau_n \leq \tau_{max}$ for some $0 < \tau_{min} \leq \tau_{max} \leq 1$ converges to the ground state v strongly in $H^1(\Omega)$.*

Proof. The proof is originally developed in [23] and we only outline its main idea here to make this paper self-contained. The key idea is to show that $u_n(x) \geq 0$ for all n by induction. Assume that $u_n \geq 0$, we will show that this implies $\mathcal{G}_{u_n} u_n \geq 0$, and with $\tau_n \leq 1$ this implies $u_{n+1} \geq 0$.

Specifically, observe that $\mathcal{G}_{u_n} u_n$ is the unique minimizer of

$$\phi(y) := (y,y)_{a_{u_n}} - 2(y,u_n)_{L^2}.$$

Since $u_n \geq 0$, we have that $\phi(|y|) \leq \phi(y)\ \forall y$. This implies that the minimizer of $\phi(\cdot)$ is nonnegative because we can always take the absolute value of the variable without increasing the functional value. Thus, $\mathcal{G}_{u_n} u_n \geq 0$. We then use the fact that u_{n+1} is the scaled weighted average of two nonnegative quantities:

$$\tilde{u}_{n+1} = (1-\tau_n)u_n + \tau_n \gamma_n \mathcal{G}_{u_n} u_n, \quad \gamma_n = \frac{(u_n,u_n)_{L^2}}{(\mathcal{G}_{u_n} u_n, \mathcal{G}_{u_n} u_n)_{a_{u_n}}} \geq 0, \quad u_{n+1} = \tilde{u}_{n+1}/\|\tilde{u}_{n+1}\|_{L^2}.$$

Thus, we establish that $u_n \geq 0$ implies $u_{n+1} \geq 0$. Since $u_0 > 0$, we have that $u_n \geq 0$ for all n.

The existence of a cluster point u^* for $\{u_n\}$ can be ensured by energy decay. This convergence to u^* is in the sense of weak convergence in $H_0^1(\Omega)$. From the above induction, u^* is nonnegative, and following an argument similar to that in Theorem 3.1 we can show that it is all positive.

Since the step size is lower-bounded, u^* must be a fixed point of $E(u)$, where grad $E(u^*) = 0$. As we mentioned above, grad $E(u^*) = 0$ implies $\nabla E(u^*) = \lambda u^*$ for some scalar λ, i.e., u^* is an eigenstate of the eigenvalue problem (1.1). From the uniqueness result of positive eigenstate in Theorem 3.1 we know that it could only be the ground state v. Therefore, $\{u_n\}$ converges to v itself.

Finally, the weak convergence in $H_0^1(\Omega)$ implies strong convergence in $L^p(\Omega)$ for $p < 6$ by the Rellich-Kondrachov embedding. This would give the convergence of energy $\{E(u_n)\}$, and consequently strong convergence in $H^1(\Omega)$. ∎

Before proceeding to the proof of Conditions (L), (D) and (S), we first need some technical lemmas.

LEMMA 3.3 (Norm equivalence). *Under Assumptions 3.1, there exist positive constants C_E, $\widetilde{C_E}$ depending only on β, M_0, V, and the domain Ω, such that*

$$C_E \|\cdot\|_{a_u} \leq \|\cdot\|_{a_0} \leq C_E^{-1} \|\cdot\|_{a_u},$$
$$\widetilde{C_E} \|\cdot\|_{a_u} \leq \|\cdot\|_{H^1} \leq \widetilde{C_E}^{-1} \|\cdot\|_{a_u}.$$

Proof. See Appendix A.1. ∎

In the next two lemmas, let λ_i and μ_i be the i-th smallest eigenvalues of $\mathcal{A}_v$ and $\mathcal{A}_u$ respectively, and v_i and w_i be their corresponding eigenfunctions satisfying $\|v_i\|_{L^2} = 1$ and $\|w_i\|_{L^2}$ (so that $v = v_1$, $\lambda = \lambda_1$). Theorem 3.1 has ensured the uniqueness of the ground state. The fact that $\mathcal{A}_v$ only has point spectrum ensures that there is a positive gap C_v between λ_1 and λ_2.

LEMMA 3.4 (Perturbation of eigenvalues and eigenfunctions). *Under Assumptions 3.1, there exists a positive constant $C = C(\beta, V, M_0, \Omega, \lambda_1, C_v)$, such that for all $u \in \mathcal{M}$ satisfying $\|u - v\|_{H^1} \leq C$, we have that $\|u - w_1\|_{L^2} \leq s$ for some $s < 1$.*

Proof. See Appendix A.2. ∎

LEMMA 3.5 (Condition (L) for the linearized operator). *Let $\mathcal{A}: X \to X$ be a symmetric and positive definite linear operator on the Hilbert space with a bounded Green's operator $\mathcal{G}$. Let μ_i denote the i-th smallest eigenvalue of $\mathcal{A}$, and w_i be its corresponding (normalized) eigenfunction. Assume that $\mu_2 > \mu_1$. Then for any u such that $\|u\|_{L^2} = 1$ and $\|u - w_1\|_{L^2} \le s < 1$, we have*

$$(u,u)_{\mathcal{A}} - (w_1,w_1)_{\mathcal{A}} \le C_L \left((u,u)_{\mathcal{A}} - \frac{1}{(u, \mathcal{G}u)_{L^2}} \right)$$

for some constant C_L that depends only on s, μ_1 and μ_2.

Proof. See Appendix A.3. ∎

Using the above technical lemmas, we are now ready to prove the following theorems. They show that the sequence $\{u_n\}$ generated by (1.3) satisfies Conditions (L), (D) and (S).

The first theorem is on Condition (L) near the ground state v of the nonlinear eigenproblem. It is the central one of the three theorems.

THEOREM 3.2. *Under Assumptions 3.1, Condition (L) is satisfied for $\|\cdot\|_X = \|\cdot\|_{a_u}$ and $\theta = \frac{1}{2}$ near the ground state v. In other words, there exists some constant $C > 0$, such that for any u in $\{u : u \in \mathcal{M}, E(u) \ge E(v), \|u - v\|_{H^1} \le C\}$, we have*

$$|E(u) - E(v)|^{\frac{1}{2}} \le C_L \|\mathrm{grad}\ E(u)\|_{a_u}.$$

Proof. First notice that for any u in the constraint set of the theorem, $E(u) - E(v) \le a_u(u,u) - a_u(v,v)$. This is because

$$\begin{aligned} E(u) - E(v) - ((u,u)_{a_u} - (v,v)_{a_u}) &= -\frac{\beta}{2}\int_\Omega u^4 - \frac{\beta}{2}\int_\Omega v^4 + \beta \int_\Omega u^2 v^2 \\ &= -\frac{\beta}{2}\int_\Omega (u^2 - v^2)^2 \le 0. \end{aligned}$$

Let w_1 be the eigenfunction corresponding to the smallest eigenvalue of $\mathcal{A}_u$, then

$$(u,u)_{a_u} - (v,v)_{a_u} \le (u,u)_{a_u} - (w_1,w_1)_{a_u}.$$

On the other hand, by (3.5), we have

$$\|\mathrm{grad}\ E(u)\|_{a_u}^2 = \left\| u - \frac{(u,u)_{L^2}}{(\mathcal{G}_u u, \mathcal{G}_u u)_{a_u}} \mathcal{G}_u u \right\|_{a_u}^2 = \left\| u - \frac{\mathcal{G}_u u}{(u, \mathcal{G}_u u)_{L^2}} \right\|_{a_u}^2 = (u,u)_{a_u} - \frac{1}{(u, \mathcal{G}_u u)_{L^2}}.$$

It suffices to show that

$$(u,u)_{a_u} - (w_1,w_1)_{a_u} \le C_L \left((u,u)_{a_u} - \frac{1}{(u, \mathcal{G}_u u)_{L^2}} \right), \tag{3.7}$$

which only involves the inner product $(\cdot,\cdot)_{a_u}$.

Using Lemma 3.4, we have that there exists $C > 0$ such that when $\|u - v\|_{H^1} < C$, we have $\|u - w_1\|_{L^2} \le s$ for some constant $s < 1$. Thus, Lemma 3.5 is applicable to $(\cdot,\cdot)_{a_u}$. This gives the above inequality on $(\cdot,\cdot)_{a_u}$, with a constant C_L depending only on $\beta, V, M_0, \Omega, \lambda_1$, and C_v. The Łojasiewicz inequality can thus be achieved. ∎

REMARK 3.2. The above proof of Condition (L) depends crucially on Lemma 3.5. Lemma 3.5 can be seen as the version of the Łojasiewicz inequality with $\theta = \frac{1}{2}$ for a

linear operator $\mathcal{A}$. So its primary consequence is the linear convergence rate of the proposed algorithm to the ground state of a linear operator $\mathcal{A}$.

The key idea of the proof of Theorem 3.2, then, is to reduce it to the inequality (3.7). The inequality (3.7) only involves the operator $\mathcal{A}_u$, which is bilinear. Although $\mathcal{A}_u$ formally depends on u, the inequality (3.7) itself is not affected by nonlinearity. So Lemma 3.5 can be applied to prove (3.7).

Thus, one way to interpret the proof of Theorem 3.2 is to view it as linearizing the nonlinear eigenproblem (1.1) using the adaptive inner product $(\cdot,\cdot)_{a_u}$, so that it preserves the Łojasiewicz property with $\theta=\frac{1}{2}$.

The next theorem is on Condition (D) for the sequence generated by the proposed algorithm.

THEOREM 3.3. *Under Assumptions 3.1, Condition (D) is satisfied for* $\|\cdot\|_X=\|\cdot\|_{a_u}$, $\|\cdot\|_Y=\|\cdot\|_{a_0}$ *if* $\{u_n\}$ *is generated by the Sobolev projected gradient descent with step size* $0<\tau_n\le\tau_{max}$ *for some* $\tau_{max}>0$, *i.e.,*

$$E(u_n)-E(u_{n+1})\ge C_D\|grad\ E(u_n)\|_{a_{u_n}}\|u_n-u_{n+1}\|_{a_0}.$$

Proof. It is obvious that $\|u_n-u_{n+1}\|_{a_0}\le\|u_n-u_{n+1}\|_{a_{u_n}}$. Since $\{u_n\}$ is generated by the Sobolev projected gradient descent algorithm, we have

$$\begin{aligned}u_{n+1}&=R(u_n-\tau_n\cdot\text{grad}\ E(u_n)),\\ \text{grad}\ E(u_n)&=u_n-\frac{(u_n,u_n)_{L^2}}{(\mathcal{G}_{u_n}u_n,\mathcal{G}_{u_n}u_n)_{a_{u_n}}}\mathcal{G}_{u_n}u_n=u_n-\frac{\mathcal{G}_{u_n}u_n}{(u_n,\mathcal{G}_{u_n}u_n)_{L^2}}.\end{aligned}$$

The second-order retraction property implies that

$$u_n-u_{n+1}=\tau_n\left(u_n-\frac{\mathcal{G}_{u_n}u_n}{(u_n,\mathcal{G}_{u_n}u_n)_{L^2}}\right)+\mathcal{O}(\tau_n^2).$$

Thus, we obtain

$$\begin{aligned}E(u_n)-E(u_{n+1})&=\left(u_n-u_{n+1},\nabla_{a_{u_n}}E(u_n)\right)_{a_{u_n}}+\mathcal{O}(\|u_n-u_{n+1}\|^2)\\&=(u_n-u_{n+1},u_n)_{a_{u_n}}+\mathcal{O}(\|u_n-u_{n+1}\|^2)\\&=\tau_n\left(u_n-\frac{\mathcal{G}_{u_n}u_n}{(u_n,\mathcal{G}_{u_n}u_n)_{L^2}},u_n\right)_{a_{u_n}}+\mathcal{O}(\tau_n^2)\\&=\tau_n\left((u_n,u_n)_{a_{u_n}}-\frac{1}{(u_n,\mathcal{G}_{u_n}u_n)_{L^2}}\right)+\mathcal{O}(\tau_n^2).\end{aligned}$$

On the other hand, we have

$$\|\text{grad}\ E(u_n)\|_{a_{u_n}}=\left((u_n,u_n)_{a_{u_n}}-\frac{1}{(u_n,\mathcal{G}_{u_n}u_n)_{L^2}}\right)^{\frac{1}{2}},$$

and

$$\begin{aligned}\|u_n-u_{n+1}\|_{a_{u_n}}&=\tau_n\left\|u_n-\frac{\mathcal{G}_{u_n}u_n}{(u_n,\mathcal{G}_{u_n}u_n)_{L^2}}\right\|_{a_{u_n}}+\mathcal{O}(\tau_n^2)\\&=\tau_n\left((u_n,u_n)_{a_{u_n}}-\frac{1}{(u_n,\mathcal{G}_{u_n}u_n)_{L^2}}\right)^{\frac{1}{2}}+\mathcal{O}(\tau_n^2).\end{aligned}$$

This implies that

$$\|\text{grad } E(u_n)\|_{a_{u_n}}\|u_n-u_{n+1}\|_{a_0}\le\tau_n\left((u_n,u_n)_{a_{u_n}}-\frac{1}{(u_n,\mathcal{G}_{u_n}u_n)_{L^2}}\right)+\mathcal{O}(\tau_n^2).$$

Therefore, there exists a $\tau_{\max}>0$ such that when $\tau\le\tau_{\max}$, there exists C_D such that Condition (D) holds. This C_D only depends on $\tau_{\max}$, but is independent of u_n. ∎

Next, we have the theorem on Condition (S) for the sequence generated by the proposed algorithm.

THEOREM 3.4. *Under Assumptions 3.1, Condition (S) is satisfied for* $\|\cdot\|_X=\|\cdot\|_{a_u}$, $\|\cdot\|_Y=\|\cdot\|_{a_0}$ *if* $\{u_n\}$ *is generated by the Sobolev projected gradient descent with step size* $0<\tau_{min}\le\tau_n\le\tau_{max}$ *for some* $0<\tau_{min}\le\tau_{max}$, *i.e.,*

$$\|u_{n+1}-u_n\|_{a_0}\ge C_S\|grad\ E(u_n)\|_{a_{u_n}}.$$

Proof. By Lemma 3.3, we have $\|u_{n+1}-u_n\|_{a_0}\ge C_E\|u_{n+1}-u_n\|_{a_{u_n}}$ for some constant C_E. Note that in the previous proof we have shown that

$$\|\text{grad } E(u_n)\|_{a_{u_n}}=\left((u_n,u_n)_{a_{u_n}}-\frac{1}{(u_n,\mathcal{G}_{u_n}u_n)_{L^2}}\right)^{\frac{1}{2}}$$

and

$$\|u_n-u_{n+1}\|_{a_{u_n}}=\tau_n\left((u_n,u_n)_{a_{u_n}}-\frac{1}{(u_n,\mathcal{G}_{u_n}u_n)_{L^2}}\right)^{\frac{1}{2}}+\mathcal{O}(\tau_n^2).$$

Therefore, when $\tau_{\min}\le\tau_n\le\tau_{\max}$ for some $0<\tau_{\min}\le\tau_{\max}$, there exists a constant C_S depending only on C_E, $\tau_{\min}$ and $\tau_{\max}$, such that

$$\|u_{n+1}-u_n\|_{a_0}\ge C_S\|\text{grad } E(u_n)\|_{a_{u_n}}.$$

∎

Finally, we deduce the following results on the exponential convergence.

THEOREM 3.5 (Convergence rate of Sobolev PGD). *If the Sobolev projected gradient descent for* $E(u)$ *converges to the ground state* v, *and the step size* $\{\tau_n\}$ *satisfies* $0<\tau_{min}\le\tau_n\le\tau_{max}$, *then it converges in the* a_0*-norm with an asymptotic exponential convergence rate.*

Proof. The proof follows directly from Theorems 2.1, 3.2, 3.3 and 3.4. ∎

COROLLARY 3.1 (Global convergence to ground state). *If the initial state* u_0 *satisfies* $u_0\ge0$ *everywhere on* Ω, *and the step size* $\{\tau_n\}$ *satisfies* $0<\tau_{min}\le\tau_n\le\tau_{max}$, *then the Sobolev projected gradient descent for* $E(u)$ *converges in the* a_0*-norm to the unique ground state with an asymptotic exponential convergence rate.*

Proof. Since the initial state is nonnegative, Lemma 3.2 ensures the strong convergence of $\{u_n\}$ to the ground state v in $H_0^1(\Omega)$. By Theorem 3.5, the asymptotic convergence rate in the a_0-norm is exponentially fast. ∎

Note that since the domain Ω is bounded, this convergence rate in the a_0-norm implies the exponential convergence rate in the H^1 or L^2 norm. We also remark that the optimal step size with theoretical guarantee depends on the values $\tau_{\min}$ and $\tau_{\max}$, which in turn depend on some properties of the ground state that is not known beforehand, but some practical choices of τ are demonstrated in the numerical experiments in Section 6.

4. Spatial discretization

To solve the eigenproblem numerically using the computational procedure in the previous sections, we need to discretize the problem in the spatial domain Ω. Let Ω_h be a spatial discretization with grid size h. Note that we only require Ω_h to be a convergent discretization, i.e., the solution to the discrete problem converges to that of the continuous problem as $h \to 0^+$, and the following analysis applies to general discretization schemes. The discretized problem can be written as

$$\min_{\|u_h\|_{L_h^2}=1,\, u_h\in\mathbb{R}^N} E_h(u_h) := \|u_h\|_{\mathcal{L}_h}^2 + \|u_h\|_{V_h}^2 + \frac{\beta}{2}\|u_h\|_{L_h^4}^4, \tag{4.1}$$

where

$$\|u_h\|_{\mathcal{L}_h}^2 = u_h^\top(-\mathcal{L}_h)u_h \cdot h^d, \quad \|u_h\|_{V_h}^2 = \sum_{i=1}^N V_h(i)u_h(i)^2 h^d, \quad \|u_h\|_{L_h^p}^p = \sum_{i=1}^N u_h(i)^p h^d.$$

Here N denotes the total number of grid points, (i) is an indexing of the grid points, i.e., $u_h(i)$ is the i-th entry of the vectorized u_h, d is the dimension of the physical space, and $\mathcal{L}_h$ is the discretized Laplacian. The linearized operator $\mathcal{A}_{u,h}$ now has a matrix representation in $\mathbb{R}^{N\times N}$:

$$\mathcal{A}_{u,h} = -\mathcal{L}_h + \operatorname{diag}\{V_h + \beta u_h^{[2]}\},$$

where $u_h^{[2]}(i) := u_h(i)^2$, i.e., $u_h^{[2]}$ is the componentwise squared vector of u_h. The respective norm is defined as $\|y\|_{\mathcal{A}_{u,h}}^2 := y^\top \mathcal{A}_{u,h} y$. We have the following results.

THEOREM 4.1 (Discrete version of Theorem 3.1). *There is a ground state v_h of the discretized problem that satisfies $v_h > 0$ everywhere on Ω_h. It is the unique positive eigenstate of (4.1). Moreover, it is both the* unique *ground state of the nonlinear eigenproblem (4.1) and the* unique *ground state of the linearized operator $\mathcal{A}_{v,h}$ up to the sign.*

Proof. The existence of the ground state follows from the compactness of the constraint set $\{u_h : u_h \in \mathbb{R}^N, \|u_h\|_{L_h^2} = 1\}$ and the boundedness of $E_h(u_h)$. Thus it suffices to prove its uniqueness and positivity. The proofs for the continuous version, i.e., Lemma 2 in [10] and Lemmas 5.3 and 5.4 in [23], need to be slightly modified to suit the discrete case. This is because the Harnack inequality and the Picone identity are only valid for continuous functions, and we need to establish their discrete counterparts.

One way to do this is to look at the convergence of the discretized eigenvector to its continuous counterpart at the small grid size limit $h \to 0^+$, see e.g. [25]. This is always possible no matter what kind of discretization we use. We do not present the details here.

Another way is to observe that the discretized Laplacian, $\mathcal{L}_h$, is an M-matrix[3] under some typical discretizations. Examples include finite difference discretization, and some P1-finite element discretizations. When $\mathcal{L}_h$ is an M-matrix, the proof can be simplified and the small h constraint can be released. In this case, the proof takes the following steps:

[3] An M-matrix is a matrix with nonnegative diagonal entries and nonpositive off-diagonal entries, with eigenvalues whose real parts are nonnegative.

(1) *For any $\mathcal{A}_{u,h}$, its eigenvector corresponding to the smallest eigenvalue can be chosen to be all positive, and is unique up to the sign.*

Since $-\mathcal{L}_h$ has positive diagonals and non-positive off-diagonals, so does $\mathcal{A}_{u,h}$. Let y be the ground state eigenvector of $\mathcal{A}_{u,h}$, then $|y|^\top \mathcal{A}_{u,h}|y| \le y^\top \mathcal{A}_{u,h} y$. This is because $\mathcal{A}_{u,h}(i,i)y(i)^2 = \mathcal{A}_{u,h}(i,i)|y(i)|^2$ for any $1 \le i \le N$, and $\mathcal{A}_{u,h}(i,j)y(i)y(j) \ge \mathcal{A}_{u,h}(i,j) \cdot |y(i)||y(j)|$ for any $i \ne j$. As y is the ground state eigenvector, this implies $y = |y|$, i.e., y is nonnegative.

We now show that y is all positive. If this is not true, then y has some positive and some zero entries. So we can always find a zero entry $y(i)$ that is spatially next to a nonzero one, say $y(j)$, i.e., $y(i) = 0$, $y(j) > 0$, and $-\mathcal{L}_h(i,j) < 0$. Then

$$\begin{aligned} 0 &= \lambda y(i) = (\mathcal{A}_{u,h} y)(i) = (-\mathcal{L}_h y)(i) + V_h(i)y(i) + \beta y(i)^3 \\ &= (-\mathcal{L}_h y)(i) = \sum_k -\mathcal{L}_h(i,k)y(k) = \sum_{k \ne i} -\mathcal{L}_h(i,k)y(k) \le -\mathcal{L}_h(i,j)y(j) < 0, \end{aligned}$$

which is a contradiction. Thus y is all positive and is unique up to the sign.

(2) *If u_h itself is the smallest eigenvector of $\mathcal{A}_{u_h,h}$, then it is also the unique global minimizer of $E_h(u)$.*

For any other $w_h \ne \pm u_h$, we have

$$\begin{aligned} E_h(w_h) - E_h(u_h) &= \|w_h\|^2_{\mathcal{A}_{u,h}} - \|u_h\|^2_{\mathcal{A}_{u,h}} + \frac{\beta}{2} \sum_{i=1}^N \left((w_h^{(i)})^4 + (u_h^{(i)})^4 - 2(w_h^{(i)})^2 (u_h^{(i)})^2 \right) h^d \\ &= \left(\|w_h\|^2_{\mathcal{A}_{u,h}} - \|u_h\|^2_{\mathcal{A}_{u,h}} \right) + \frac{\beta}{2} \sum_{i=1}^N \left((w_h^{(i)})^2 - (u_h^{(i)})^2 \right)^2 h^d > 0. \end{aligned}$$

Thus u_h is the unique global minimizer of $E_h(u)$.

(3) *There is a unique positive eigenstate of (4.1), which is the ground state of (4.1) and the ground state of the linearized operator.*

Any positive iteration sequence stays positive with gradient descent iteration. The compactness of the constraint set ensures the existence of a sub-sequential limit point v_h, which is nonnegative. The fact that v_h is the minimizer of $E_h(u)$ implies that it is an eigenstate of $\mathcal{A}_{v,h}$. By Step (1), this eigenstate is all positive and is thus the smallest eigenstate of $\mathcal{A}_{v,h}$. By Step (2), it is also the unique global minimizer of $E_h(u)$.

∎

THEOREM 4.2 (Discrete version of Theorem 3.5). *If the Sobolev PGD for $E_h(u)$ converges to the ground state v_h, and the step size $\{\tau_n\}$ satisfies $0 < \tau_{min} \le \tau_n \le \tau_{max}$, then it converges with an asymptotic exponential convergence rate.*

Proof. Theorem 4.1 ensures that v_h is still the "double" ground state of both $E_h(u)$ and $\mathcal{A}_{v_h,h}$. Thus, Theorems 3.2, 3.3, and 3.4 can all be generalized to the discretized case in the same way. The exponential convergence rate follows from the master theorem 2.1. ∎

COROLLARY 4.1 (Discrete version of Corollary 3.1). *If the initial state u_0 satisfies $u_0(i) \ge 0$ $\forall i$, and the step size $\{\tau_n\}$ satisfies $0 < \tau_{min} \le \tau_n \le \tau_{max}$, then the Sobolev PGD for $E_h(u)$ converges to the unique ground state v_h with an asymptotic exponential convergence rate.*

Proof. The proof follows similarly from the nonnegativity and uniqueness results of Theorem 4.1 and the exponential convergence result of Theorem 4.2. ∎

5. Generalization to other nonlinear eigenproblems

The Sobolev PGD points out a new direction for first-order fast solvers of nonlinear eigenproblems and higher (than quadratic) order optimization problems. Its application is thus well beyond the Gross-Pitaevskii eigenvalue problem. The operator class and the form of the objective function can be generalized. For example, consider

$$-\Delta v + Vv + \beta |v|^{2\alpha} v = \lambda v \tag{5.1}$$

for general $\alpha > 0$. This ground state equation and the corresponding time-dependent nonlinear Schrödinger equation are locally well-posed in $H^1(\mathbb{R}^d)$ as long as $2\alpha + 2 < \frac{2d}{\max\{d-2,0^+\}}$, see e.g. [18] and references therein. The previous Gross-Pitaevskii eigenvalue problem corresponds to the case $\alpha = 1$.

In general, Theorem 3.5 holds true for any $\alpha > 0$. The adaptive inner product remains well-posed and the ground state remains a "double" eigenstate. The change of inner product from $a_v(\cdot)$ to $a_u(\cdot)$ in the proof of Theorem 3.2 essentially relies on the convexity of the last term $\int |\cdot|^{2\alpha+2}$ in the energy functional $E(\cdot)$. Therefore, extensions of the previous results in both spatially continuous and discretized cases are easy. We do not present the details here.

It is also common in physics that the diffusion is not homogeneous in all spatial directions. For example, it can be stronger in two physical directions and weaker in the third one. More generally, we have

$$-\nabla \cdot (A(x)\nabla v) + Vv + \beta |v|^{2\alpha} v = \lambda v \tag{5.2}$$

where the coefficient $A(x) \in L^\infty(\Omega)^{d\times d}$, $A(x)$ is symmetric and coercive. An interesting discrete counterpart to this is the nonlinear Schrödinger equation on metric trees (e.g. [15]), where the Laplacian is replaced by a graph Laplacian on a tree-graph $\mathcal{G}$.

When restricted to a bounded domain, so that the lowest part of the spectrum is always point spectrum, our previous arguments still hold. In the elliptic case, the discretized $\mathcal{A}_h$ may or may not be an M-matrix, but one can always turn to the small grid size limit $h \to 0^+$ limit when necessary.

For an even broader class of nonlinear eigenproblems or constrained optimization problems, the Sobolev gradient descent may still be applicable, but it is not clear whether *exponential convergence* is still true. It can be seen from previous sections that the convergence rate relies on the (L) condition, which in turn relies on the ground state v being the ground state of the linearized operator $\mathcal{A}_v$ at v, i.e., the so-called *"double ground state"* property. This is a nontrivial property in general, although it can be true for some operators like the biharmonic operator under certain conditions.

We discuss here one specific generalization of nonlinear Schrödinger eigenproblem, and demonstrate that the Sobolev PGD indeed has the potential of tackling previously formidable problems. The problem of interest is

$$-\Delta v + Vv + \beta |v|^2 v - \delta \Delta(|v|^2) v = \lambda v, \tag{5.3}$$

where $\delta \geq 0$. In other words, we add a higher-order interaction term $-\delta\Delta(|v|^2)$ to the Gross-Pitaevskii problem. The corresponding energy functional is

$$E(u) = \int |\nabla u|^2 + V|u|^2 + \frac{\beta}{2}|u|^4 + \frac{\delta}{2}\left|\nabla |u|^2\right|^2. \tag{5.4}$$

The above eigenproblem and its variational form are analyzed in [5]. Moreover, in [7] the authors propose to minimize the energy functional (5.4) by reformulating it as $E(\rho)=\int|\nabla\sqrt{\rho}|^2+V\rho+\frac{\beta}{2}\rho^2+\frac{\delta}{2}|\nabla\rho|^2$, where $\rho:=|u|^2$. This reformulation facilitates the minimization, but it also suffers from the lack of continuity of $|\nabla\sqrt{\rho}|$ near $\rho\to 0^+$. This has to be treated with extra care, and a regularization term has to be added, which complicates the analysis. Therefore, instead of replacing $|u|^2$ with ρ, we propose to minimize $E(u)$ with respect to u directly with the Sobolev PGD.

Assume that Assumptions 3.1 still hold. Define the manifold $\mathcal{M}$ with an extra constraint:

$$\mathcal{M}:=\left\{z\in H_0^1(\Omega):\|u\|_{L^2}=1,\|u\|_{L^\infty}\le M_0,\|\nabla u\|_{L^\infty}\le M_1\right\}.$$

Define the adaptive linearized operator and the respective inner products as follows:

$$\begin{aligned}
&(z,w)_{a_u}:=\int_\Omega \nabla z\nabla w+Vzw+\beta u^2zw+\delta\nabla(uz)\nabla(uw),\\
&(z,\mathcal{A}_u w)_{L^2}:=(z,w)_{a_u},\\
&(z,w)_{a_0}:=\int_\Omega \nabla z\nabla w+Vzw,\qquad \forall z,w\in\mathcal{T}_u\mathcal{M},\quad u\in\mathcal{M}.
\end{aligned}$$

Then we have the following results.

LEMMA 5.1. *The ground state v of (5.3) satisfies $v>0$ everywhere on Ω. It is the unique positive eigenstate of (5.3). It is also both the unique ground state of (5.3) and that of the linearized operator $\mathcal{A}_v$ up to the sign.*

Proof. Following the same arguments as in Lemma 2 in [8], the extended $E(u)$ as in (5.4) still admits a nonnegative minimizer v. According to [5, Theorem 2.2], we know that $v\in C^{1,1}(\bar\Omega)$. This implies that v, $\nabla v\in L^\infty(\Omega)$. Thus, the nonnegative v can still be made positive by the Harnack inequality. Also, the linearized operator $\mathcal{A}_v$ still has a unique positive ground state, which is exactly the above v. Thus the "double ground state" property remains true.

We now show that (5.3) has a unique positive eigenstate by a contradiction argument. Suppose instead that there is a different positive eigenstate $\tilde v>0$ with its eigenvalue $\tilde\lambda$, and $E(\tilde v)>E(v)$. Using the Picone identity, $\int\nabla\tilde v\nabla(\frac{v^2}{\tilde v})\le\int(\nabla v)^2$, we have

$$\begin{aligned}
&\tilde\lambda-\lambda=\tilde\lambda(v,v)_{L^2}-(v,v)_{a_v}=\tilde\lambda\left(\tilde v,\frac{v^2}{\tilde v}\right)_{L^2}-(v,v)_{a_v}=\left(\tilde v,\frac{v^2}{\tilde v}\right)_{a_{\tilde v}}-(v,v)_{a_v}\\
&=\int\nabla\tilde v\cdot\nabla\left(\frac{v^2}{\tilde v}\right)+Vv^2+\beta\tilde v^2v^2+\delta\nabla(\tilde v^2)\nabla(v^2)-\int(\nabla v)^2+Vv^2+\beta v^4+\delta(\nabla(v^2))^2\\
&\le\int(\nabla v)^2+Vv^2+\frac{\beta}{2}(v^4+\tilde v^4)+\frac{\delta}{2}\left((\nabla(v^2))^2+(\nabla(\tilde v^2))^2\right)-\int(\nabla v)^2+Vv^2+\beta v^4+\delta(\nabla(v^2))^2\\
&=\int\frac{\beta}{2}\tilde v^4+\frac{\delta}{2}(\nabla(\tilde v^2))^2-\int\frac{\beta}{2}v^4+\frac{\delta}{2}(\nabla(v^2))^2=(\tilde\lambda-E(\tilde v))-(\lambda-E(v)),
\end{aligned}$$

i.e.,

$$E(\tilde v)\le E(v).$$

This contradicts our assumption that $E(\tilde v)>E(v)$. ∎

The next lemma shows that the eigenvalue and eigenfunction perturbation results stated in Lemma 3.4 hold similarly for (5.3).

LEMMA 5.2. *Let λ_i and μ_i be the i-th smallest eigenvalues of $\mathcal{A}_v$ and $\mathcal{A}_u$ respectively, and v_i and w_i be their corresponding eigenvectors (so that $v=v_1$). Let $C_v := \lambda_2 - \lambda_1$ denote the eigenvalue gap. Then there exists a positive constant $C = C(\beta,\delta,V,M_0,M_1,\Omega,\lambda_1,C_v)$, such that for all $\|u-v\|_{H^1} < C$, $u \in \mathcal{M}$, we have $\|u - w_1\|_{L^2} \le s$ for some $s<1$.*

Proof. See Appendix A.4. ∎

THEOREM 5.1. *If the initial state satisfies $u_0 \ge 0$ everywhere on Ω, then $\{u_n\}_{n=0}^{\infty}$ generated by the Sobolev PGD with step size $0 < \tau_{min} \le \tau_n \le \tau_{max}$ converges to the unique ground state v of (5.3) with an asymptotic exponential convergence rate.*

Proof. First, the Sobolev PGD sequence starting from a positive initial value remains positive as before, and convexity ensures convergence to a nonnegative local minimizer of $E(u)$, which must also be the global minimizer and the ground state of (5.3). This convergence can be proved to be a strong H^1 convergence by the Sobolev embedding and the convergence of energy.

In order to establish exponential convergence, it suffices to show that Conditions (L), (D) and (S) all hold for $\{u_n\}_{n=0}^{\infty}$. The nonnegativity of δ ensures the equivalence of a_0 and a_u norms. Thus Conditions (D) and (S) hold. Condition (L) follows from Lemma 5.2 and Lemma 3.5. ∎

The above results establish the exponential convergence of the Sobolev PGD for problem (5.3) for any $\delta \ge 0$. Numerical evidence shows that the Sobolev PGD for this problem converges very well just as the original Gross-Pitaevskii eigenproblem. This is a demonstration that the Sobolev gradient descent has the potential to be generalized to study some continuous or discrete high degree optimization problems. We believe that this method has the potential to be extended to a broader class of problems as long as certain assumptions are satisfied, which is left for our future work.

6. Numerical experiments

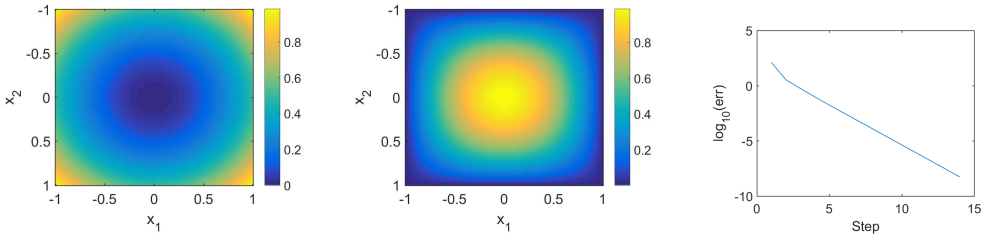
In this section, we demonstrate the convergence of the Sobolev PGD method using some numerical examples. We show that exponential convergence rate is attained both for the original eigenproblem (1.1) and for its extension (5.3). We also observe and discuss some interesting phenomena that one may encounter in numerical experiments.

6.1. Gross-Pitaevskii eigenproblem in 2D. We first look at the Gross-Pitaevskii eigenproblem (1.1) in two dimensions. Let the domain be $\Omega = [-1,1]^2 \subset \mathbb{R}^2$ with Dirichlet boundary condition. The problems are discretized with P1 Lagrange finite element method. The grid is a uniform grid with fixed size $h = 2 \cdot 2^{-8}$ throughout this section.

The first example is a single well potential $V(x) = \frac{1}{2}|x|^2$. It is well known that the Anderson localization [3] is present in this setting. We set $\beta = 1$. The initial guess z_0 is chosen as the eigenvector corresponding to the smallest eigenvalue of $\mathcal{A}_0$. It is strictly positive over the whole domain Ω. The step size is $\tau = 1$.

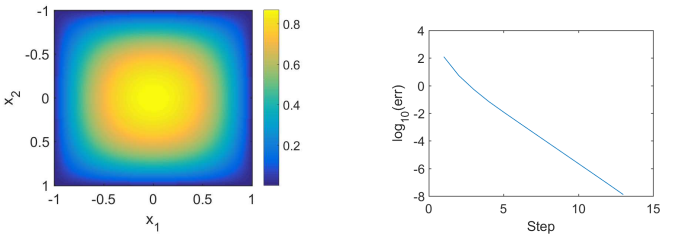
Figure 6.1a shows the profile of the potential $V(x)$. Figure 6.1b is the profile of the computed ground state with $\beta = 1$. Figure 6.1c displays the log H^1-error convergence $\log_{10}(\|u_n - v\|_{H^1}/\|v\|_{H^1})$. It can be seen that the Sobolev PGD converges in just a few steps with an exponential (linear) convergence rate.

By increasing β, there is a greater nonlinearity in the problem. When $\beta \gg 1$, the quartic term $\frac{\beta}{2}|u|^4$ would dominate the energy functional (1.2). This would be a significant barrier to some traditional methods. Yet the Sobolev PGD remains stable and fast. Figures 6.2a to 6.2d show the log H^1-error convergence and the profiles of the

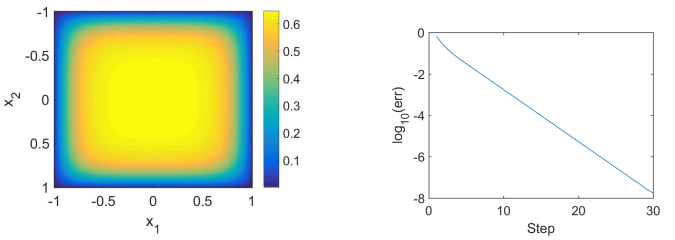


(a) Single well potential $V(x) = \frac{1}{2}|x|^2$ (b) Ground state when $\beta = 1$ (c) Log H^1-error convergence

Fig. 6.1: *Example of (1.1) with single well potential* $V = \frac{1}{2}|x|^2$ *and* $\beta = 1$.



(a) Ground state when $\beta = 10$ (b) Convergence when $\beta = 10$



(c) Ground state when $\beta = 100$ (d) Convergence when $\beta = 100$

Fig. 6.2: *Example of (1.1) with single well potential* $V = \frac{1}{2}|x|^2$ *and* $\beta = 10$ *or* 100.

respective ground states with $\beta = 10$ and $\beta = 100$ respectively. With the Sobolev PGD, there is only a mild increase in the computational complexity, and the iteration still converges exponentially fast as predicted.

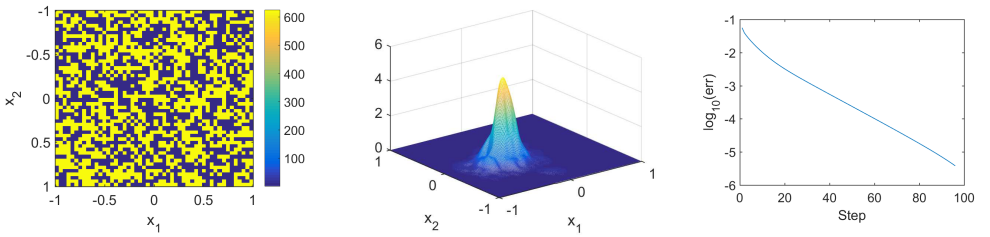
6.2. Localization under the disordered potential. The second example is a disordered potential V. Its fully discrete counterpart, the randomized potential on the lattice $\mathbb{Z}^d$, has been extensively studied for its rich behaviour in spectral gaps, exponential localization of eigenstates near the bottom of the spectrum, and implications about the "mobility edge" conjecture in quantum physics and random matrix theory [14, 20].

In our semi-lattice example, the localization of the ground state is also present. In the experiment, $V(x)$ is generated as follows. The extent of disorder is determined by a parameter $K = 50$. This means that the domain Ω is divided into $K \times K$ cells. The

value of V(x) in each cell is either 1 or $1/K^2$, randomly chosen with equal probability.

Figure 6.3a shows the profile of $V(x)$. Figure 6.3b displays the computed ground state with $\beta = 0.5$. It can be seen that the ground state is concentrated in a small region whose diameter is about a few times the interaction length of the disorder. Figure 6.3c shows the convergence rate of the Sobolev PGD iteration for this example.

To facilitate convergence, we have chosen $\tau = 1.5$. Although Corollary 3.1 requires a small τ, in the numerical experiments we find that choosing $\tau > 1$ results in significantly faster convergence. This is in accordance with the empirical findings of [23].



(a) Profile of the disordered potential (b) Profile of the ground state (c) Log H^1-error convergence

Fig. 6.3: *Example of (1.1) with a disordered potential and* $\beta = 0.5$.

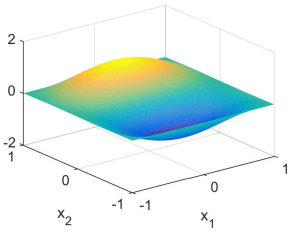
6.3. Asymptotic escape of Sobolev PGD from saddle states. It is interesting to look at the asymptotic behaviour of the Sobolev gradient descent method if starting from a non-positive initial value. Recall that Corollary 3.1 only ensures exponential convergence to the global ground state from $u_0 \geq 0$. When this condition is violated, it is a priori unknown what the iteration will converge to. It is possible that there are other spurious fixed points, including local minimizers and saddle points. The first-order condition ensures that all these spurious fixed points are eigenstates. But the convergence rate to such points is unknown.

As for the spatially discretized case, the Hilbert manifold $\mathcal{M}$ becomes a Riemannian manifold, and the spectra of the operators become finite. As is proved in [24] and references therein, a random initialization almost surely avoids saddles and converges only to local or global minimizers. It means that if an excited state is a strict saddle point, then a random initialization is very unlikely to converge to that state. As for the spatially continuous case, it is reasonable to expect the same phenomenon, although the analysis could be more difficult due to the infinite dimension of $\mathcal{M}$ and the infinite number of eigenstates.

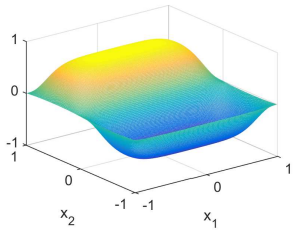
In the subsequent numerical experiments, we let $V(x) = \frac{1}{2}|x|^2$ and $\beta = 100$. We will use an example to show that for an excited state that is a strict saddle, it has a very thin converging set close to measure zero. Thus, using Sobolev PGD to compute excited states could be unstable. The accuracy of the computed excited states could be limited.

First, we let the initialization u_0 be the second-smallest eigenvector of $\mathcal{A}_0$. This u_0 is positive on half of Ω and negative on the other half. It is displayed in Figure 6.4a. We then let Sobolev PGD iterate a few steps. Figure 6.4b displays the computed state u^* when the algorithm stops. Figure 6.4c shows the decrease of the log L^2 error with respect to u^*. We also compute the manifold Hessian at u^* and find that it has at least one negative eigenvalue. Thus u^* is a strict saddle state.

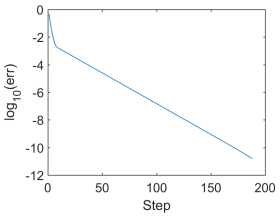
Next, we add a small perturbation to u_0: we let $\hat{u}_0 = u_0 + \epsilon \cdot \eta$, where η is a random noise that is of the same order as u_0, and the parameter ϵ controls the magnitude of



(a) Profile of initial state u_0

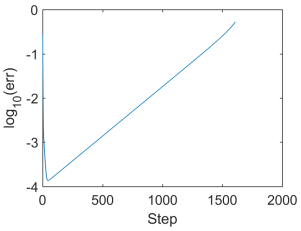


(b) Profile of computed state u^*

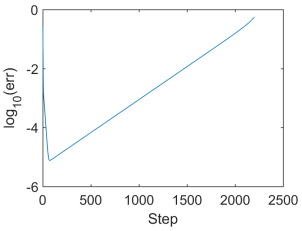


(c) Log H^1-error convergence

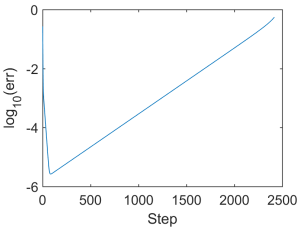
Fig. 6.4: *Behavior of Sobolev PGD for (1.1) with* $u_0 \not\geq 0$.



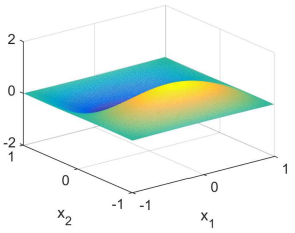
(a) Noise level $\epsilon = 10^{-2}$



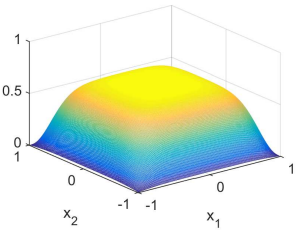
(b) Noise level $\epsilon = 10^{-3}$



(c) Noise level $\epsilon = 10^{-4}$



(d) Profile of $\hat{u}_0 = u_0 + 10^{-4}\eta$



(e) Profile of computed state starting from $\hat{u}_0$

Fig. 6.5: *Asymptotic escape from saddle state under small perturbations. Figures (a)-(c) displays the distances to the saddle state* u^* *starting from* $\hat{u}_0 = u_0 + \epsilon \cdot \eta$.

noise. We let Sobolev PGD start from $\hat{u}_0$ and trace its evolution. What we observe is that, as long as there is a small perturbation, Sobolev PGD escapes from the previous saddle state and converges to the ground state. The parameter ϵ can be chosen as small as 10^{-4} and this effect is still present.

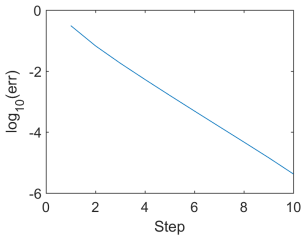
Specifically, Figures 6.5a to 6.5c demonstrate the evolution of the log-distance to the precomputed closest excited state u^*. We choose $\epsilon = 10^{-2}$, 10^{-3}, and 10^{-4}, respectively. Saddle escape behavior can be observed in all three cases. We can see that the distance to the excited state first goes down, then goes up. Figure 6.5e shows the computed state starting from $\hat{u}_0$, and it is the ground state.

In general, first-order optimization methods, including Sobolev PGD as well as other methods in the gradient descent family, are not good choices for the computation of excited states. They rely on a good enough initialization (like the above u_0 without noise) and could suffer from numerical instability issues. One has to resort to other

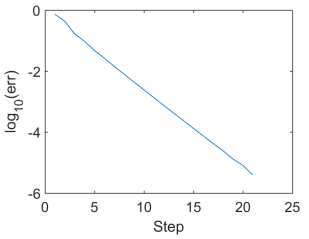
methods if the goal is to obtain high accuracy. We will explore this topic in our upcoming work.

6.4. High order interaction. We now look at Problem (5.3) with an extra high order interaction term. This adds additional nonlinearity to the problem. Consider the same domain $\Omega = [-1,1]^2 \subset \mathbb{R}^2$ and spatial discretization size $h = 2 \cdot 2^{-8}$. Let $V(x) = \frac{1}{2}|x|^2$ still be the single well potential. The first example is $\beta = 10$ and $\delta = 1$. Figure 6.6a shows the log error convergence. The iteration converges in a few steps and shows a good convergence rate.

In the second example, we increase δ and look at the problem with strong high order interaction. We choose $\beta = 100$ and $\delta = 100$. Figure 6.6b shows the log error convergence. The convergence rate is slower but stable.



(a) Convergence when $\beta = 10$, $\delta = 1$



(b) Convergence when $\beta = 100$, $\delta = 100$

Fig. 6.6: *Examples of (5.3) with different nonlinear effects*

7. Conclusion

In this paper, we analyzed the exponential convergence of the a_u-Sobolev gradient descent method without resorting to the time-continuous gradient flow. To this purpose, we introduced a general convergence tool using the Łojasiewicz inequality, and adapted it to the setting of infinite dimensional Hilbert manifold and mixed norms. By proving the (L), (D) and (S) conditions for the Sobolev PGD, we were able to unveil the mechanism behind the good performance of the Sobolev PGD for the Gross-Pitaevskii eigenproblem (1.1), which was only empirically observed in previous works.

The success of the Sobolev PGD on the Gross-Pitaevskii eigenproblem inspires us to further explore alternative fast solvers for more general nonlinear eigenproblems and optimizations with high degree objective functions. Our analysis revealed that the essential condition is the "double ground state" property, namely the ground state of the nonlinear problem is also the unique ground state of the linearized operator at that point. This can be rigorously proved in some cases and seems to be true in a number of physical applications of interest based on empirical evidence. Specifically, we showed that this condition is satisfied for a nonlinear Schrödinger eigenproblem with extra high order interaction term. Thus the Sobolev PGD works well for this problem and has superiority over previous methods.

Acknowledgements. This research was in part supported by NSF grants DMS-1912654 and DMS-1907977. The author would like to thank Thomas Y. Hou for the helpful comments on earlier versions of this work, and Zhenzhen Li for introducing the Łojasiewicz inequality to the author. The author would also like to acknowledge the warm hospitality of Oberwolfach Research Institute for Mathematics during the seminar *Beyond Numerical Homogenization*, where the early ideas of this work started.

Appendix. Proofs of technical lemmas.

A.1. Proof of Lemma 3.3.

Proof. As for the equivalence between $\|\cdot\|_{a_0}$ and $\|\cdot\|_{a_u}$, the second part of the inequality holds for all $0<C_E\le 1$ since u^2 is nonnegative. For the first part, by Poincaré inequality, $\|z\|_{L^2}^2\le C_P|z|_{H^1}^2$ for some domain constant $C_P=C_P(\Omega)$. Thus, we have

$$\begin{aligned}\|z\|_{a_0}^2-C_E\|z\|_{a_u}^2&=(1-C_E)|z|_{H^1}^2+\int_\Omega((1-C_E)V-C_E\beta u^2)z^2\\&\ge(1-C_E)|z|_{H^1}^2-C_E\beta\int_\Omega u^2z^2\\&\ge(1-C_E-C_E\beta M_0^2C_P)|z|_{H^1}^2,\qquad\forall z\in H_0^1(\Omega),\quad C_E\le 1.\end{aligned}$$

Take $0<C_E\le 1/(1+\beta M_0^2C_P)$, then $C_E\|z\|_{a_u}^2\le\|z\|_{a_0}^2$.

As for the equivalence between $\|\cdot\|_{a_u}$ and $\|\cdot\|_{H^1}$, we have

$$\begin{aligned}\|z\|_{H^1}^2-\widetilde{C_E}\|z\|_{a_u}^2&=\|z\|_{H^1}^2-\widetilde{C_E}|z|_{H^1}^2-\widetilde{C_E}\int_\Omega(V+\beta u^2)z^2\\&\ge\left(1-\widetilde{C_E}-\widetilde{C_E}C_P(\|V\|_{L^\infty}+\beta M_0^2)\right)|z|_{H^1}^2,\qquad\forall z\in H_0^1(\Omega),\quad\widetilde{C_E}\le 1.\end{aligned}$$

Take $0<\widetilde{C_E}\le 1/(1+C_P(\|V\|_{L^\infty}+\beta M_0^2))$, then $\widetilde{C_E}\|z\|_{a_u}^2\le\|z\|_{H^1}^2$. On the other hand,

$$\begin{aligned}\widetilde{C_E}^{-1}\|z\|_{a_u}^2-\|z\|_{H^1}^2&=(\widetilde{C_E}^{-1}-1)|z|_{H^1}^2+\widetilde{C_E}^{-1}\int_\Omega(V+\beta u^2)z^2-\|z\|_{L^2}\\&\ge\left(C_P^{-1}(\widetilde{C_E}^{-1}-1)+\widetilde{C_E}^{-1}\beta M_0^2-1\right)\|z\|_{L^2}.\end{aligned}$$

Take $0<\widetilde{C_E}\le(1+C_P\beta M_0^2)/(1+C_P)$, then $\|z\|_{H^1}^2\le\widetilde{C_E}^{-1}\|z\|_{a_u}^2$. The final choice of $\widetilde{C_E}$ is the smaller of the two. ∎

A.2. Proof of Lemma 3.4.

Proof. For notational simplicity, we allow the constants C, C' to change their meanings through the proof. We also denote

$$t:=\|u-v\|_{H^1}.$$

Using the variational form of the eigenvalues, we have

$$\begin{aligned}\mu_1&=\min_{\substack{z\in H_0^1(\Omega),\\ \|z\|_{L^2}=1}}(z,z)_{a_u}\le(v,v)_{a_u},\\ \lambda_1&=\min_{\substack{z\in H_0^1(\Omega),\\ \|z\|_{L^2}=1}}(z,z)_{a_v}\le(w_1,w_1)_{a_v},\\ \lambda_1+\lambda_2&=\min_{\substack{z_1,z_2\in H_0^1(\Omega),\\ \|z_1\|_{L^2}=\|z_2\|_{L^2}=1,\\ z_1\perp z_2}}(z_1,z_1)_{a_v}+(z_2,z_2)_{a_v}\le(w_1,w_1)_{a_v}+(w_2,w_2)_{a_v}.\end{aligned}$$

We will use the above relations to bound the gap between μ_1 and λ_1, and λ_2 and μ_2. First, we have

$$\mu_1\le(v,v)_{a_u}=(v,v)_{a_v}+\int_\Omega\beta(u^2v^2-v^4)$$

$$
\begin{aligned}
&= \lambda_1 + \int_\Omega \beta v^2 (u+v)(u-v) \\
&\le \lambda_1 + 2\beta M_0^3 \int_\Omega |u-v| \\
&\le \lambda_1 + C(\beta, M_0, \Omega) \cdot t.
\end{aligned}
$$

Therefore, there exists $C = C(\beta, M_0, \Omega)$ such that when $t \le C$,

$$
\mu_1 \le \lambda_1 + \frac{1}{6} C_v. \tag{A.1}
$$

Next, we note that

$$
\begin{aligned}
\lambda_1 + \lambda_2 &\le (w_1, w_1)_{a_v} + (w_2, w_2)_{a_v} \\
&= (w_1, w_1)_{a_u} + (w_2, w_2)_{a_u} + \int_\Omega \beta (v^2 - u^2)(w_1^2 + w_2^2) \\
&= \mu_1 + \mu_2 + \int_\Omega \beta (v+u)(v-u)(w_1^2 + w_2^2).
\end{aligned} \tag{A.2}
$$

To estimate $\|w_1\|_{L^\infty}$, note that it is the weak solution of

$$
-\Delta w_1 + V w_1 + \beta u^2 w_1 = \mu_1 w_1.
$$

Since $V, u \in L^\infty(\Omega)$, by elliptic regularity, we get

$$
\begin{aligned}
\|w_1\|_{H^2} &\le C(\beta, V, M_0, \Omega)(\|w_1\|_{H^1} + \mu_1 \|w_1\|_{L^2}) \\
&\le C(\beta, V, M_0, \Omega) + C'(\beta, V, M_0, \Omega) \cdot \mu_1.
\end{aligned}
$$

When $d \le 3$, using Sobolev embedding, we obtain

$$
\|w_1\|_{L^\infty} \le C(\Omega) \|w_1\|_{H^2}.
$$

Since we have shown that $\mu_1 \le \lambda_1 + C \cdot t$, putting them together we have

$$
\|w_1\|_{L^\infty} \le C(\beta, V, M_0, \Omega, \lambda_1) + C'(\beta, V, M_0, \Omega, \lambda_1) \cdot t.
$$

Similarly, we can prove that[4]

$$
\|w_2\|_{L^\infty} \le C(\beta, V, M_0, \Omega, \lambda_1, \lambda_2) + C'(\beta, V, M_0, \Omega, \lambda_1, \lambda_2) \cdot t.
$$

Plugging them back into (A.2), we have

$$
(w_1, w_1)_{a_v} + (w_2, w_2)_{a_v} \le \mu_1 + \mu_2 + \left(C(\beta, V, M_0, \Omega, \lambda_1, \lambda_2) + C'(\beta, V, M_0, \Omega, \lambda_1, \lambda_2) \cdot t\right)^2 \cdot t.
$$

Therefore, there exists $C = C(\beta, V, M_0, \Omega, \lambda_1, \lambda_2)$, such that when $t \le C$,

$$
\lambda_1 + \lambda_2 \le \mu_1 + \mu_2 + \frac{1}{6} C_v. \tag{A.3}
$$

Combining (A.1) and (A.3), we have

$$
\mu_1 \le \lambda_1 + \frac{1}{6} C_v, \qquad \mu_2 \ge \lambda_2 - \frac{1}{3} C_v, \qquad \mu_2 - \mu_1 \ge \frac{1}{2} C_v. \tag{A.4}
$$

[4]We omit the details of showing $\mu_2 \le \lambda_2 + C \cdot t$ by showing $\mu_1 + \mu_2 \le \lambda_1 + \lambda_2 + C \cdot t$ using the variational form.

Next, note that

$$\begin{aligned}\lambda_1 \le (w_1,w_1)_{a_v} &= (w_1,w_1)_{a_u} + \int_\Omega \beta(v^2-u^2)w_1^2 \\ &\le \mu_1 + C(\beta,V,M_0,\Omega)\|w_0\|_{L^\infty}^2 \cdot t \\ &\le \mu_1 + (C(\beta,V,M_0,\Omega) + C'(\beta,V,M_0,\Omega)\cdot t)^2 \cdot t.\end{aligned}$$

Therefore, there exists $C = C(\beta,V,M_0,\Omega,\lambda_1)$ such that when $t \le C$,

$$\lambda_1 \le \mu_1 + \frac{1}{6}C_v. \tag{A.5}$$

Equations (A.1), (A.4) and (A.5) contain all the relations between $\lambda_1, \lambda_2, \mu_1$, and μ_2 that we will need.

Since $\{w_i\}_{i=1}^\infty$ forms an orthonormal basis of $H_0^1(\Omega)$, in order to estimate $\|u-w_1\|_{L^2}$, it suffices to bound $(u,u)_{a_u} - \mu_1$. Note that

$$\begin{aligned}(u,u)_{a_u} - \lambda_1 &= (u,u)_{a_u} - (v,v)_{a_v} \\ &= (u,u)_{a_u} - (v,v)_{a_u} + \int_\Omega \beta(u^2v^2 - v^4) \\ &\le (\|u\|_{a_u} + \|v\|_{a_u}) \cdot \|u-v\|_{a_u} + \int_\Omega \beta v^2(u+v)(u-v) \\ &\le C(\beta,V,M_0,\Omega)(\|u\|_{H^1} + \|v\|_{H^1}) \cdot \|u-v\|_{H^1} + \int_\Omega \beta v^2(u+v)(u-v) \\ &\le C(\beta,V,M_0,\Omega)\cdot t.\end{aligned}$$

The fourth inequality uses the norm equivalence in Lemma 3.3. Thus, there exists $C = C(\beta,V,M_0,\Omega)$, such that when $t \le C$,

$$(u,u)_{a_u} - \lambda_1 \le \frac{1}{12}C_v. \tag{A.6}$$

Combining (A.4), (A.5) and (A.6), we have

$$(u,u)_{a_u} - \mu_1 \le \frac{1}{4}C_v \le \frac{1}{2}(\mu_2 - \mu_1).$$

Assume that $u = \sum_{i=1}^\infty c_i w_i$, where $\sum_{i=1}^\infty c_i^2 = 1$. Then we get

$$(u,u)_{a_u} - \mu_1 = \sum_{i=1}^\infty c_i^2\mu_i - \mu_1 \ge c_1^2\mu_1 + \sum_{i=2}^\infty c_i^2\mu_2 - \mu_1 = (1-c_1^2)(\mu_2-\mu_1).$$

Since $(u,u)_{a_u} - \mu_1 \le \frac{1}{2}(\mu_2-\mu_1)$, we have

$$1 - c_1^2 \le \frac{1}{2}, \qquad |c_1| \ge \frac{1}{\sqrt{2}}.$$

If $c_1 \le -1/\sqrt{2}$, we can use $-w_1$ to replace w_1. Thus, we always have $c_1 \ge 1/\sqrt{2}$. This gives

$$\|u-w_1\|_{L^2} = \sqrt{2-2c_1} \le \sqrt{2-\sqrt{2}} < 1.$$

In other words, $s \le \sqrt{2-\sqrt{2}}$. The constant C in the statement of the lemma is the smallest of all the constants C, C' in the proof. Since $\lambda_2 = \lambda_1 + C_v$, the dependence on λ_2 is the dependence on C_v. ∎

A.3. Proof of Lemma 3.5.

Proof. Since μ_2 is strictly greater than μ_1, we can split $\mathcal{A}$ and u as

$$\mathcal{A}=\mathcal{A}^{(1)}+\mathcal{A}^{(2)},\quad \mathcal{A}^{(1)}=\mathcal{A}P_{w_1},\quad \mathcal{A}^{(2)}=\mathcal{A}P_{w_1}^{\perp},$$
$$u=u^{(1)}+u^{(2)},\quad u^{(1)}=P_{w_1}u,\quad u^{(2)}=P_{w_1}^{\perp}u.$$

Here P_{w_1} is the orthogonal projection onto the subspace of w_1 under the L^2 inner product, and $P_{w_1}^{\perp}=id-P_{w_1}$. Then $\mathcal{A}^{(1)}u^{(1)}=\mu_1u^{(1)}$, and $(u^{(2)},u^{(2)})_{\mathcal{A}^{(2)}}\geq\mu_2\|u^{(2)}\|_{L^2}^2$ since $u^{(2)}\perp w_1$. By definition of $\mathcal{G}$, $(u,\mathcal{G}v)_{\mathcal{A}}=(u,v)_{L^2}$ for any $u,v\in X$. We have

$$\begin{aligned}
&(u,\mathcal{G}u^{(1)})_{L^2}=\mu_1^{-1}\|u^{(1)}\|_{L^2}^2,\\
&(u,\mathcal{G}u^{(2)})_{L^2}=(u^{(1)},\mathcal{G}u^{(2)})_{L^2}+(u^{(2)},\mathcal{G}u^{(2)})_{L^2}=(u^{(2)},\mathcal{G}u^{(2)})_{L^2},\\
&(u^{(2)},\mathcal{G}u^{(2)})_{L^2}=(\mathcal{G}u^{(2)},\mathcal{G}u^{(2)})_{\mathcal{A}}\geq\mu_2\|\mathcal{G}u^{(2)}\|_{L^2}^2\\
&\qquad=\mu_2\|u^{(2)}\|_{L^2}^{-2}\cdot(\|\mathcal{G}u^{(2)}\|_{L^2}^2\|u^{(2)}\|_{L^2}^2)\geq\mu_2\|u^{(2)}\|_{L^2}^{-2}\cdot(u^{(2)},\mathcal{G}u^{(2)})_{L^2}^2,\\
&\qquad\text{i.e., }(u,\mathcal{G}u^{(2)})_{L^2}\leq\mu_2^{-1}\|u^{(2)}\|_{L^2}^2.
\end{aligned}$$

Therefore, the objective inequality is transformed into

$$\begin{aligned}
&C_L\left((u,u)_{\mathcal{A}}-\frac{1}{(u,\mathcal{G}u)_{L^2}}\right)-((u,u)_{\mathcal{A}}-(w_1,w_1)_{\mathcal{A}})\\
&=(C_L-1)(u,u)_{\mathcal{A}}-\frac{C_L}{(u,\mathcal{G}u)_{L^2}}+\mu_1\\
&=(C_L-1)((u^{(1)},u^{(1)})_{\mathcal{A}^{(1)}}+(u^{(2)},u^{(2)})_{\mathcal{A}^{(2)}})-\frac{C_L}{(u,\mathcal{G}u^{(1)})_{L^2}+(u,\mathcal{G}u^{(2)})_{L^2}}+\mu_1\\
&\geq(C_L-1)(\mu_1\|u^{(1)}\|_{L^2}^2+\mu_2\|u^{(2)}\|_{L^2}^2)-\frac{C_L}{\mu_1^{-1}\|u^{(1)}\|_{L^2}^2+\mu_2^{-1}\|u^{(2)}\|_{L^2}^2}+\mu_1\\
&=(C_L-1)(\mu_1+(\mu_2-\mu_1)\|u^{(2)}\|_{L^2}^2)-\frac{C_L\mu_1\mu_2}{\mu_2+(\mu_1-\mu_2)\|u^{(2)}\|_{L^2}^2}+\mu_1\\
&=(\mu_2-\mu_1)\frac{((C_L-1)\mu_2-C_L\mu_1)\|u^{(2)}\|_{L^2}^2-(C_L-1)(\mu_2-\mu_1)\|u^{(2)}\|_{L^2}^4}{\mu_2+(\mu_1-\mu_2)\|u^{(2)}\|_{L^2}^2}.
\end{aligned}$$

We look for C_L and u such that the above is greater than or equal to 0. In fact, for any $C_L>1$, if

$$0\leq\|u^{(2)}\|_{L^2}^2\leq\frac{(C_L-1)\mu_2-C_L\mu_1}{(C_L-1)(\mu_2-\mu_1)},$$

then this is satisfied. Note that $\|u-v_1\|_{L^2}\leq s$ implies $\|u^{(2)}\|_{L^2}^2\leq s^2$. So the requirement on C_L is

$$C_L\geq1+\frac{\mu_2}{(\mu_2-\mu_1)(1-s^2)}.$$

∎

A.4. Proof of Lemma 5.2.

Proof. The main idea of the proof is the same as that of Lemma 3.4 so we only point out their differences here. For example, to estimate $\mu_1-\lambda_1$, we have

$$\mu_1\leq(v,v)_{a_u}=(v,v)_{a_v}+\int_\Omega\beta(u^2v^2-v^4)+\int_\Omega\delta\left((\nabla(uv)^2-\nabla(v^2)^2)\right)$$

$$= \lambda_1 + \int_\Omega \beta v^2 (u+v)(u-v) + \int_\Omega \delta(\nabla(uv) + \nabla(v^2))(\nabla(uv) - \nabla(v^2)).$$

The second term is bounded in the same way as the proof of Lemma 3.4. Only the third term containing high-order interaction is new. To bound the third term, we note that

$$\begin{aligned}
&\int_\Omega \delta(\nabla(uv) + \nabla(v^2))(\nabla(uv) - \nabla(v^2)) \\
&= \delta \int_\Omega (v\nabla u + u\nabla v + 2v\nabla v)(v\nabla u + u\nabla v - 2v\nabla v) \\
&\le 4\delta M_0 M_1 \int_\Omega |v\nabla u + u\nabla v - 2v\nabla v| \\
&= 4\delta M_0 M_1 \int_\Omega |v(\nabla u - \nabla v) + (u-v)\nabla v| \\
&\le C(\delta, M_0, M_1, \Omega)\|u-v\|_{H^1}.
\end{aligned}$$

Similar bounds can be obtained in the estimation of $(\lambda_1+\lambda_2)-(\mu_1+\mu_2)$, $\lambda_1-\mu_1$, and $(u,u)_{a_u}-\mu_1$. The dependence of the constant C only has two additional dependencies which are δ and M_1. ∎

REFERENCES

[1] P.A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, 2008. 1

[2] M.H. Anderson, J.R. Ensher, M.R. Matthews, C.E. Wieman, and E.A. Cornell, *Observation of Bose-Einstein condensation in a dilute atomic vapor*, Science, 269(5221):198–201, 1995. 1

[3] P.W. Anderson, *Absence of diffusion in certain random lattices*, Phys. Rev., 109(5):1492, 1958. 6.1

[4] W. Bao and Y. Cai, *Mathematical theory and numerical methods for Bose-Einstein condensation*, Kinet. Relat. Models, 6(1):1–135, 2013. 1

[5] W. Bao, Y. Cai, and X. Ruan, *Ground states of Bose–Einstein condensates with higher order interaction*, Phys. D, 386:38–48, 2019. 1, 5, 5

[6] W. Bao and Q. Du, *Computing the ground state solution of Bose–Einstein condensates by a normalized gradient flow*, SIAM J. Sci. Comput., 25(5):1674–1697, 2004. 1

[7] W. Bao and X. Ruan, *Computing ground states of Bose-Einstein condensates with higher order interaction via a regularized density function formulation*, SIAM J. Sci. Comput., 41(6):B1284–B1309, 2019. 1, 5

[8] E. Cancès, R. Chakir, and Y. Maday, *Numerical analysis of nonlinear eigenvalue problems*, J. Sci. Comput., 45(1-3):90–117, 2010. 1, 3.2, 5

[9] E. Cancès, G. Kemlin, and A. Levitt, *Convergence analysis of direct minimization and self-consistent iterations*, SIAM J. Matrix Anal. Appl., 42(1):243–274, 2021. 1

[10] E. Cancès and C. Le Bris, *On the convergence of SCF algorithms for the Hartree-Fock equations*, ESAIM: Math. Model. Numer. Anal., 34(4):749–774, 2000. 1, 4

[11] Y. Chi, Y.M. Lu, and Y. Chen, *Nonconvex optimization meets low-rank matrix factorization: An overview*, IEEE Trans. Signal Process., 67(20):5239–5269, 2019. 1

[12] 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(5):2447–2467, 2010. 1

[13] I. Danaila and B. Protas, *Computation of ground states of the Gross–Pitaevskii functional via Riemannian optimization*, SIAM J. Sci. Comput., 39(6):B1102–B1129, 2017. 1

[14] P. Deift, *Some open problems in random matrix theory and the theory of integrable systems. II*, Symmetry Integr. Geom., 13:016, 2017. 6.2

[15] S. Dovetta, E. Serra, and P. Tilli, *NLS ground states on metric trees: existence results and open questions*, J. London Math. Soc., 102(3):1223–1240, 2020. 5

[16] G. Dusson and Y. Maday, *A posteriori analysis of a nonlinear Gross–Pitaevskii-type eigenvalue problem*, IMA J. Numer. Anal., 37(1):94–137, 2017. 1

[17] E. Faou and T. Jézéquel, *Convergence of a normalized gradient algorithm for computing ground states*, IMA J. Numer. Anal., 38(1):360–376, 2018. 1

[18] R.L. Frank, *Ground states of semi-linear PDEs*, Lecture notes, 2014. 5
[19] P. Frankel, G. Garrigos, and J. Peypouquet, *Splitting methods with variable metric for Kurdyka–Łojasiewicz functions and general convergence rates*, J. Optim. Theory Appl., 165(3):874–900, 2015. 1, 2
[20] J. Fröhlich and T. Spencer, *Absence of diffusion in the Anderson tight binding model for large disorder or low energy*, Commun. Math. Phys., 88(2):151–184, 1983. 6.2
[21] D. Gilbarg and N.S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Springer, 2015. 3.2
[22] P. Henning, A. Målqvist, and D. Peterseim, *Two-level discretization techniques for ground state computations of Bose-Einstein condensates*, SIAM J. Numer. Anal., 52(4):1525–1550, 2014. 1
[23] P. Henning and D. Peterseim, *Sobolev gradient flow for the Gross–Pitaevskii eigenvalue problem: Global convergence and computational efficiency*, SIAM J. Numer. Anal., 58(3):1744–1772, 2020. 1, 1, 3.1, 3.2, 3.2, 4, 6.2
[24] T.Y. Hou, Z. Li, and Z. Zhang, *Analysis of asymptotic escape of strict saddle sets in manifold optimization*, SIAM J. Math. Data Sci., 2(3):840–871, 2019. 6.3
[25] J.R. Kuttler, *Finite difference approximations for eigenvalues of uniformly elliptic operators*, SIAM J. Numer. Anal., 7(2):206–232, 1970. 4
[26] E.H. Lieb, R. Seiringer, and J. Yngvason, *Bosons in a trap: A rigorous derivation of the Gross-Pitaevskii energy functional*, in W. Thirring (eds.), The Stability of Matter: From Atoms to Stars, Springer, 685–697, 2001. 1
[27] R. Schneider and A. Uschmajew, *Convergence results for projected line-search methods on varieties of low-rank matrices via Łojasiewicz inequality*, SIAM J. Optim., 25(1):622–646, 2015. 1, 2
[28] H. Zhang, A. Milzarek, Z. Wen, and W. Yin, *On the geometric analysis of a quartic-quadratic optimization problem under a spherical constraint*, Math. Program., 21, 2021. 1