Normalized Wolfe-Powell-type local minimax method for finding multiple unstable solutions of nonlinear elliptic PDEs

Wei Liu¹,†, Ziqing Xie¹ & Wenfan Yi²,*

¹Key Laboratory of Computing and Stochastic Mathematics (Ministry of Education), Hunan Normal University, Changsha 410081, China;
²School of Mathematics, Hunan University, Changsha 410082, China
Email: wliu@m.scnu.edu.cn, ziqingxie@hunnu.edu.cn, wfyi@hnu.edu.cn

Received August 11, 2021; accepted January 17, 2023; published online April 21, 2023

Abstract The local minimax method (LMM) proposed by Li and Zhou (2001, 2002) is an efficient method to solve nonlinear elliptic partial differential equations (PDEs) with certain variational structures for multiple solutions. The steepest descent direction and the Armijo-type step-size search rules are adopted in Li and Zhou (2002) and play a significant role in the performance and convergence analysis of traditional LMMs. In this paper, a new algorithm framework of the LMMs is established based on general descent directions and two normalized (strong) Wolfe-Powell-type step-size search rules. The corresponding algorithm framework, named the normalized Wolfe-Powell-type LMM (NWP-LMM), is introduced with its feasibility and global convergence rigorously justified for general descent directions. As a special case, the global convergence of the NWP-LMM combined with the preconditioned steepest descent (PSD) directions is also verified. Consequently, it extends the framework of traditional LMMs. In addition, conjugate-gradient-type (CG-type) descent directions are utilized to speed up the NWP-LMM. Finally, extensive numerical results for several semilinear elliptic PDEs are reported to profile their multiple unstable solutions and compared with different algorithms in the LMM’s family to indicate the effectiveness and robustness of our algorithms. In practice, the NWP-LMM combined with the CG-type direction performs much better than its known LMM companions.

Keywords semilinear elliptic PDE, multiple unstable solution, local minimax method, normalized strong Wolfe-Powell-type search rule, conjugate-gradient-type descent direction, general descent direction, global convergence

MSC(2020) 35J20, 35B38, 65N12, 65J15, 65J30

Citation Liu W, Xie Z Q, Yi W F. Normalized Wolfe-Powell-type local minimax method for finding multiple unstable solutions of nonlinear elliptic PDEs. Sci China Math, 2023, 66: 2361–2384, https://doi.org/10.1007/s11425-021-2093-1

† Current address: South China Research Center for Applied Mathematics and Interdisciplinary Studies, South China Normal University, Guangzhou 510631, China
* Corresponding author

© Science China Press 2023
1 Introduction

Various nonlinear problems in physics, chemistry, biology and materials science can be reduced to consider multiple solutions of the Euler-Lagrange equation associated with a continuously Fréchet-differentiable nonlinear functional $E$ defined on a real Hilbert space $H$, i.e.,

$$E'(u) = 0, \quad u \in H,$$  \hspace{1cm} (1.1)

where $E'$ is the Fréchet-derivative of $E$. Actually, the solutions of the Euler-Lagrange equation (1.1) are called critical points of the functional $E$. The most well-known candidates for critical points are local extrema, to which classical variational and optimization methods have contributed a lot.

With the development of new experimental techniques, it has been possible to observe local unstable equilibria or transient excited states in numerous physical/chemical/biological systems. Consequently, their theoretical and numerical studies have attracted increasing attention \cite{2, 5, 16, 25, 32, 38}. However, these local unstable equilibria or transient excited states are related to critical points that are not local extrema and are called saddle points. Virtually, in terms of the instability analysis, for a critical point $u_*$ with its second-order Fréchet-derivative $E''(u_*)$ existing, its instability can be depicted by its Morse index (MI) \cite{2}, which is defined as the maximal dimension of subspaces of $H$ where the linear operator $E''(u_*)$ is negative-definite. In fact, for a nondegenerate critical point $u_*$, i.e., $E''(u_*)$ is invertible if its MI is equal to 0, it is a strict local minimizer and then a stable critical point, while if its MI is greater than 0, then it is an unstable critical point. Generally speaking, a higher MI means more instability.

Owing to the instability and multiplicity of saddle points, we see that the design and analysis of numerical methods for grasping saddle points in a stable way are much more challenging than those for stable critical points. In recent years, various numerical methods have been developed to capture saddle points in a stable way, such as the (climbing) string method \cite{10, 26}, the gentlest ascent dynamics \cite{11} and the (shrinking) dimer method \cite{15, 37}. Nevertheless, the methods mentioned above mainly focus on finding saddle points with MI = 1.

In this paper, we are interested in stable and efficient numerical computations for multiple saddle points with high MIs. Existing methods in this area include the search extension method \cite{3, 4} and its modified versions \cite{21, 30}, the augmented partial Newton method and its variants \cite{19, 31}, and the high-index optimization-based shrinking dimer method \cite{36} and its extension to non-gradient systems \cite{35}. Recently, some dynamical methods for finding constrained saddle points with high MIs were also developed \cite{22, 34}. On the other hand, motivated by classical minimax theorems in the critical point theory (see, e.g., \cite{25} and the references therein) and numerical research of Choi and McKenna \cite{6}, Ding et al. \cite{9} and Chen et al. \cite{5}, Li and Zhou \cite{17} proposed a local minimax method (LMM) for various high-MI saddle points based on a local minimax characterization of them. Then in \cite{32}, Xie et al. modified the LMM with a significant relaxation on the domain of the local peak selection, which is a crucial notion for the LMM and will be illustrated in detail later. According to \cite{17, 32}, the LMM grasps a saddle point with MI = $n$ ($n \in \mathbb{N}^+$) by dealing with a two-level local minimax problem as

$$\min_{v \in S_H} \max_{w \in [L, v]} E(w),$$  \hspace{1cm} (1.2)

where $S_H = \{v \in H : \|v\| = 1\}$ is the unit sphere with $\| \cdot \|$ the norm in $H$, $L \subset H$ is a given $(n - 1)$-dimensional closed subspace usually constructed based on some known or previously found critical points, and $[L, v] = \{tv + w_L : t \geq 0, w_L \in L\}$ denotes a closed half subspace. Actually, the inner local maximization is an optimization problem in the $n$-dimensional half subspace $[L, v]$, which can be solved efficiently by standard optimization algorithms in Euclidean spaces. The outer constrained local minimization, which is generally infinite-dimensional and much more challenging in numerical computations, is the major concern of the LMM. For the sake of handling this task, the LMM adopts a normalized iterative scheme (NIS) with the steepest descent direction $d_k^{SD} = -g_k$, i.e.,

$$v_{k+1} = v_k(\alpha_k) = \frac{v_k - \alpha_k g_k}{\|v_k - \alpha_k g_k\|}, \quad w_{k+1} = p(v_{k+1}), \quad k = 0, 1, 2, \ldots,$$  \hspace{1cm} (1.3)
where $\alpha_k > 0$ is a step-size and $g_k = \nabla E(w_k) \in H$ is the gradient of $E$ at $w_k = p(v_k)$ with $p(v_k)$ representing a local maximizer of $E$ on $[L, v_k]$, known as the so-called local peak selection of $E$ with respect to $E$ at $v_k$. Actually, the local peak selection $p(v_k)$ can be expressed as $p(v_k) = I_k v_k + w_k^L$ for some $t_k \geq 0$ and $w_k^L \in L$ [17].

One of the fundamental problems for the NIS (1.3) is choosing a suitable step-size $\alpha_k$. In the earliest implementations of the LMM, a normalized exact step-size search rule was employed and aimed to find the step-size $\alpha_k > 0$ such that the functional $E(p(v_k(\alpha)))$ attains its minimum [17], i.e.,

$$E(p(v_k(\alpha_k))) = \min_{\alpha > 0} E(p(v_k(\alpha))), \quad k = 0, 1, 2, \ldots \quad (1.4)$$

Nevertheless, such a step-size search rule is very expensive in practical computations, and it is even hard to establish the global convergence of the corresponding LMM. To compensate for this shortage, several normalized inexact step-size search rules with the advantages of low computational cost and easy implementation have been introduced in the literature to choose the step-size $\alpha_k > 0$ such that the decrease amount $E(p(v_k)) - E(p(v_k(\alpha_k))) > 0$ is acceptable.

Note that a widely applied normalized inexact step-size search rule in traditional LMMs is the normalized Armijo-type step-size search rule, which was first introduced in [18] and further simplified in [33] as the following form:

$$E(p(v_k(\alpha_k))) \leq E(p(v_k)) - \frac{1}{4} \alpha_k t_k \|g_k\|^2, \quad k = 0, 1, 2, \ldots \quad (1.5)$$

In fact, the factor $1/4$ can be replaced by any constant $\sigma \in (0, 1)$ (see [20]). However, the decrease condition (1.5) is satisfied for all the sufficiently small step-sizes (see Figure 1), and hence some artificial safeguards are needed to prevent step-sizes from being too small and the algorithm from interminable backtracking [18,39]. Actually, the backtracking strategy chooses the largest step-size in the sequence $\{\lambda \rho^m\}_{m \in \mathbb{N}}$ (for a given trial step-size $\lambda > 0$ and a backtracking factor $\rho \in (0, 1)$) that satisfies the normalized Armijo-type search rule. It plays an important role not only in the numerical implementation but also in the convergence analysis of the NA-LMM. Nevertheless, a choice of appropriate parameters $\lambda$ and $\rho$ is not known a priori. Recently, a normalized Goldstein-type step-size search rule was proposed in [20] to guarantee a sufficient decrease of the functional and prevent step-sizes from being too small simultaneously. The feasibility and the global convergence analysis of the normalized Goldstein-type LMM (NG-LMM) were also provided in [20]. Actually, the normalized Goldstein-type step-size search rule makes progress with two inequalities, which can be formulated as

$$-\sigma \alpha_k t_k \|g_k\|^2 \leq E(p(v_k(\alpha_k))) - E(p(v_k)) \leq -\sigma \alpha_k t_k \|g_k\|^2, \quad k = 0, 1, 2, \ldots \quad (1.6)$$

with constants $\sigma$ and $\delta$ satisfying $0 < \sigma < \delta < 1$. Unfortunately, as shown in Figure 1, the normalized Goldstein-type step-size search rule may exclude the minimizer $\alpha_*$ of $E(p(v_k(\alpha)))$ outside the acceptable interval $[\bar{\alpha}_1, \bar{\alpha}_2]$. Thus, more effective and reasonable step-size search rules may be devoted to the LMM for credibly capturing saddle points of the functional $E$.

Furthermore, the convergence rate is another fundamental problem of the NIS (1.3) for solving the outer minimization in the two-level local optimization problem (1.2). As shown in the NIS (1.3), the steepest descent direction $d_k^S = -g_k = -\nabla E(p(v_k))$ has been chosen as a descent direction in all the existing LMMs since 2001. As a result, they have certain limitations in terms of the convergence rate. It is well known that in line search algorithms for unconstrained optimizations in Euclidean spaces, many choices of descent directions may have better performance than the steepest descent direction, such as the conjugate gradient (CG) direction and the quasi-Newton direction, which usually have rapid convergence rates. Therefore, in this paper, we try to design some improved iterative schemes of the form as (1.3) by replacing the steepest descent direction $d_k^S = -g_k$ by other more efficient descent directions $d_k \in H$ for the outer minimization process to
improve the numerical performance and convergence rate of the LMM’s family. We note that the CG and quasi-Newton methods in the optimization theory are often used in combination with some Wolfe-Powell line search strategy [24, 28, 29] that consists of the Armijo condition and a curvature condition. In fact, such a curvature condition on the step-size is particularly important for both the algorithm implementation and the convergence analysis of the CG and quasi-Newton methods.

Inspired by the discussions above, this paper aims to develop a new LMM framework based on general descent directions and the (strong) Wolfe-Powell-type step-size search rules, called a normalized Wolfe-Powell-type LMM (NWP-LMM), to capture multiple unstable solutions of the Euler-Lagrange equation (1.1) and provide the possibility to speed up the convergence. By employing some curvature properties, two types of normalized Wolfe-Powell-type step-size search rules will be introduced for the LMM with general descent directions. Their mathematical justifications and global convergence will be established rigorously for general descent directions by making full use of these curvature properties, which obviously distinguish from those of the NA-LMM [18, 32] and the NG-LMM [20]. Finally, two types of descent directions, i.e., the preconditioned steepest descent (PSD) direction and the CG-type descent direction, will be proposed and compared to be implemented in our NWP-LMM for computing multiple unstable solutions of several semilinear elliptic partial differential equations (PDEs), such as the nonlinear Schrödinger equation (NLSE), the Hénon equation and the Chandrasekhar equation. Indeed, it will be seen that the CG-type descent direction can greatly speed up our NWP-LMM. It is worthwhile to point out that the steepest descent direction can be replaced by a general descent direction in the device of the traditional normalized Armijo-type and Goldstein-type LMMs. Furthermore, both their feasibility and global convergence can be verified in the lines of our approach.

The rest of this paper is organized as follows. Firstly, some preliminaries for the LMM are provided in Section 2. Then in Section 3, the NWP-LMM framework based on general descent directions and the normalized (strong) Wolfe-Powell-type step-size search rules is introduced. Its feasibility and some related properties are also discussed in this section. Global convergence of the NWP-LMM with general descent directions is verified rigorously in Section 4. In addition, two different types of descent directions, i.e., the PSD and CG-type descent directions, are proposed and analyzed in Section 5 to feasibly implement our NWP-LMM. Furthermore, Section 6 reports the detailed numerical results in 2 dimensions including the numerical comparison of different LMMs for the above-mentioned semilinear elliptic PDEs to illustrate the effectiveness and robustness of our approach. Finally, some conclusions are drawn in Section 7.

![Figure 1](image.png)

Figure 1  Illustration of normalized Armijo- and Goldstein-type step-size search rules in traditional LMMs: the acceptable interval of the Armijo-type step-size is $(0, \bar{\alpha}_2]$, while the acceptable interval of the Goldstein-type step-size is $[\bar{\alpha}_1, \bar{\alpha}_2]$
2 Preliminaries

For the convenience of discussions later, we introduce some notations and basic lemmas for the LMM in this section.

Let $(\cdot, \cdot)$ and $\| \cdot \|$ be, respectively, the inner product and the norm in the Hilbert space $H$. Denote by $2^H$ the set of all the subsets of the Hilbert space $H$, by $S_H = \{ v \in H : \|v\| = 1 \}$ the unit sphere in $H$ and by $X^\perp$ the orthogonal complement to a subspace $X \subset H$. Suppose that $L$, serving as a so-called support space later, is a given closed finite-dimensional subspace in $H$. Define the half subspace $[L, v] := \{ tv + w^L : t \geq 0, w^L \in L \}$ for any $v \in S_H$. Throughout this paper, we assume that the functional $E$ has a local minimizer at 0 $\in H$ and focus on finding nontrivial saddle points of $E$. The local peak selection, a crucial notion, is defined as follows (see [17, 20, 32]).

**Definition 2.1.** The peak mapping of $E$ with respect to $L$ is a set-valued mapping $P : S_H \to 2^H$ such that for any $v \in S_H$, $P(v)$ is the set of all the local maximum points of $E$ on $[L, v]$. A peak selection of $E$ with respect to $L$ is a single-valued mapping $p : S_H \to H$ such that

$$p(v) \in P(v), \quad \forall v \in S_H.$$  

For a given $v \in S_H$, we say that $E$ has a local peak selection with respect to $L$ at $v$ if there are a neighborhood $N(v)$ of $v$ and a mapping $p : N(v) \cap S_H \to H$ such that

$$p(u) \in P(u), \quad \forall u \in N(v) \cap S_H.$$  

Let $(\cdot, \cdot)$ denote the duality pairing between $H$ and its dual space $H^*$. By definition, each local peak selection $p(v)$, with $v \in S_H$, can be expressed as $p(v) = t_v v + w_v^L$ with $t_v \geq 0$ and $w_v^L \in L$. To avoid the degeneracy, we always assume that $t_v > 0$, i.e., $p(v) \notin L$, in the subsequent analysis. The above definition implies that $p(v)$ belongs to the well-known Nehari manifold $N_E := \{ u \in H \setminus \{0\} : E'(u), u = 0 \}$, which contains all the nontrivial critical points of the functional $E$. In fact, the following orthogonality holds obviously.

**Lemma 2.2** (See [20, 32]). Assume that $E \in C^1(H, \mathbb{R})$ has a local peak selection $p$ with respect to $L$ at $v \in S_H$ satisfying $p(v) \notin L$. Then $(E'(p(v)), w) = 0$, $\forall w \in [L, v]$. In particular, $(E'(p(v)), p(v)) = 0$.

The following property follows from direct computations and is frequently utilized in the feasibility and convergence discussions for LMM-type algorithms.

**Lemma 2.3** (See [20]). Let $\bar{v} \in S_H \setminus L$. Suppose that the local peak selection $p$ of $E$ with respect to $L$ is continuous at $\bar{v}$. Define $p(v) = t_v v + w_v^L$ and $p(\bar{v}) = t_{\bar{v}} \bar{v} + w_{\bar{v}}^L$, where $t_v, t_{\bar{v}} \geq 0$ and $w_v^L, w_{\bar{v}}^L \in L$. If $v \to \bar{v}$, then $t_v \to t_{\bar{v}}$ and $w_v^L \to w_{\bar{v}}^L$.

Imitating the similar lines of the proof of [17, Theorem 2.1], we obtain the following local minimax principle and refer to [20, Theorem 4.1].

**Theorem 2.4.** If $E \in C^1(H, \mathbb{R})$ has a local peak selection with respect to $L$ at $\bar{v} \in S_H \setminus L$, denoted by $p(\bar{v}) = t_{\bar{v}} \bar{v} + w_{\bar{v}}^L$, satisfying (i) $p$ is continuous at $\bar{v}$, (ii) $t_{\bar{v}} > 0$, and (iii) $\bar{v}$ is a local minimizer of $E(p(v))$ on $S_H$, then $p(\bar{v}) \notin L$ is a critical point of $E$.

Since the local peak selection $p(\bar{v})$ is a local maximizer of $E$ on the half subspace $[L, \bar{v}]$, Theorem 2.4 characterizes a saddle point of $E$ as a solution of the local minimax problem (1.2), or equivalently, the local minimization problem of $E$ on the solution submanifold

$$M = \{ p(v) : v \in S_H \}.$$  

When a saddle point $u^*$ of $E$ (known as an unstable critical point in $H$) can be characterized as a local solution of the two-level optimization problem (1.2) of the form $u^* = p(v^*)$ with $v^*$ minimizing $E(p(v))$ on $S_H$, it becomes stable on $M$, i.e.,

$$E(u^*) = \min_{u \in M} E(u). \quad (2.1)$$
We remark here that under certain conditions, similar to [17, Theorem 2.2] and [18, Theorem 1.5], the existence of the local minimizer of \( E(p(v)) \) on a subspace of \( S_H \) can be verified by employing the Ekeland’s variational principle. In addition, Theorem 2.4 also indicates an important feature of the LMM, which shows that it can stably find different saddle points and avoid computing those we have found. In fact, in order to obtain multiple saddle points, the LMM needs to repeatedly solve the local minimization problem (2.1) or the two-level local minimax problem (1.2) for different choices of \( L \), which is usually spanned by some found critical points. Under assumptions of Theorem 2.4, a local minimizer of \( E \) on \( M \) is a critical point different from those in \( L \). As a result, Theorem 2.4 provides a mathematical justification that the LMM can find unstable saddle points of the functional \( E \) in a stable way.

Virtually, suitable descent algorithms can work for the minimization process in the optimization problem (1.2) or (2.1). As noted above, in traditional LMMs, the steepest descent direction serves as a search direction to numerically solve the local minimax problem (1.2), while in this paper, the NWP-LMM will be constructed and analyzed for general descent directions.

### 3 Normalized Wolfe-Powell-type LMM

In this section, an NWP-LMM framework for general descent directions will be proposed to capture multiple saddle points of the functional \( E \) via solving the optimization problem (1.2). In order to achieve this goal, two types of normalized Wolfe-Powell-type step-size search rules with general descent directions will be introduced and analyzed by involving curvature conditions. We adopt the same notations as those in Section 2 unless specified and begin with some essential properties.

Let \( v \in S_H \setminus L \) and \( d \in [L, v]^\perp \). Write \( v = v^L + v^\perp \) with \( v^L \in L \) and \( v^\perp \in L^\perp \setminus \{0\} \). For all \( \alpha \in \mathbb{R} \), the following orthogonal decomposition holds (see Figure 2):

\[
v(\alpha) = \frac{v + \alpha d}{\|v + \alpha d\|} = \frac{v + \alpha d}{\sqrt{1 + \alpha^2\|d\|^2}} = v^L(\alpha) + v^\perp(\alpha) \in S_H,
\]

where

\[
v^L(\alpha) = \frac{v^L}{\sqrt{1 + \alpha^2\|d\|^2}} \in L, \quad v^\perp(\alpha) = \frac{v^\perp + \alpha d}{\sqrt{1 + \alpha^2\|d\|^2}} \in L^\perp.
\]

Obviously, for all \( \alpha \in \mathbb{R} \), \( \|v^L(\alpha)\| \leq \|v^L\| \). Since \( \|v(\alpha)\|^2 = \|v^L(\alpha)\|^2 + \|v^\perp(\alpha)\|^2 = 1 \), it follows that

\[
\|v^\perp(\alpha)\| \geq \|v^\perp\| > 0, \quad \forall \alpha \in \mathbb{R}.
\]

Consequently, \( v(\alpha) \in S_H \setminus L, \forall \alpha \in \mathbb{R} \). Furthermore, we have the following property.

![Figure 2](image-url) (Color online) Illustration of the normalized iterative scheme (3.1)
Lemma 3.1. Let $v \in S_H \setminus L$, $d \in [L, v]^+$ and $v(\alpha)$ be expressed as in (3.1). Then

$$
\|v(\alpha + s) - v(\alpha)\| \leq |s|\|d\|, \quad \forall \alpha, s \in \mathbb{R}.
$$

(3.2)

Proof. Define $\ell_1 = \sqrt{1 + (\alpha + s^2)\|d\|^2}$ and $\ell_2 = \sqrt{1 + \alpha^2\|d\|^2}$. Noting that $\ell_1, \ell_2 \geq 1$, we have

$$
\begin{align*}
\ell_1^2 - \ell_2^2 &= (\ell_1 - \ell_2)(\ell_1 + \ell_2) \\
&= (\ell_1 - \ell_2)(\ell_1 + (\alpha + s)\|d\|) \\
&= (\ell_1 - \ell_2)(\ell_1 + (\alpha + s)\|d\|) \\
&= (\ell_1 - \ell_2)(\ell_1 + (\alpha + s)\|d\|).
\end{align*}
$$

This completes the proof. \hfill \square

Set $p$ as a local peak selection of $E$ with respect to $L$ at $v \in S_H \setminus L$. The following lemma is crucial in constructing the normalized Wolfe-Powell-type step-size search rules.

Lemma 3.2. Suppose $E \in C^1(H, \mathbb{R})$ and let $v \in S_H \setminus L$, $d \in [L, v]^+$, $v(\alpha)$ be expressed as in (3.1) and $p(\alpha) = t_v(\alpha)\langle v(\alpha), v(\alpha) \rangle$ be a local peak selection of $E$ with respect to $L$ at $v(\alpha)$. If $p$ is locally Lipschitz continuous around $v(\alpha)$ and $t_v(\alpha) > 0$, then the composite function $\alpha \rightarrow E(p(\alpha))$ is continuously differentiable and

$$
\frac{d}{d\alpha} E(p(\alpha)) = \hat{i}_v(\alpha)(E'(p(\alpha)), d),
$$

(3.3)

where $\hat{i}_v(\alpha) = t_v(\alpha)/\sqrt{1 + \alpha^2\|d\|^2}$.

Proof. The mean value theorem states that when $s \in \mathbb{R}$ is close to zero,

$$
E(p(v(\alpha + s))) - E(p(v(\alpha))) = \langle E'(p(\alpha)), p(v(\alpha + s)) - p(v(\alpha)) \rangle \\
= \langle E'(v(\alpha)), p(v(\alpha + s)) - p(v(\alpha)) \rangle \\
\leq \|E'(v(\alpha))\|_H \|p(v(\alpha + s)) - p(v(\alpha))\|,
$$

(3.4)

where $\xi = p(v(\alpha)) + \theta(p(v(\alpha + s)) - p(v(\alpha)))$ for some $\theta = \theta(\alpha) \in (0, 1)$. By the continuity of $E'$, $p$ and $v(\alpha)$, we have $\|E'(\xi) - E'(p(\alpha)))\|_H$, $\rightarrow 0$ as $s \rightarrow 0$. The local Lipschitz continuity of $p$ and Lemma 3.1 imply that the right-hand side of (3.4) is $o(||p(v(\alpha + s)) - p(v(\alpha))||) = o(||v(\alpha + s) - v(\alpha)||) = o(s||d||)$, and therefore, by Lemma 2.2,

$$
\begin{align*}
E(p(v(\alpha + s))) - E(p(v(\alpha))) &= \langle E'(p(\alpha)), p(v(\alpha + s)) - p(v(\alpha)) \rangle + o(s||d||) \\
&= t_v(\alpha + s)\langle E'(p(\alpha)), v(\alpha + s) + o(s||d||) \\
&= \frac{t_v(\alpha + s)\|d\|}{\sqrt{1 + (\alpha + s)^2||d||^2}}(E'(p(\alpha))), d + o(s||d||).
\end{align*}
$$

From Lemma 2.3, we obtain $t_v(\alpha + s) \rightarrow t_v(\alpha)$ as $s \rightarrow 0$. It follows that

$$
\lim_{s \rightarrow 0} \frac{E(p(v(\alpha + s))) - E(p(v(\alpha)))}{s} = \frac{t_v(\alpha)}{\sqrt{1 + \alpha^2\|d\|^2}}(E'(p(\alpha))) d, \quad \alpha \in \mathbb{R}.
$$

In other words, $E(p(\alpha)))$ is differentiable with respect to $\alpha$ and (3.3) holds. Finally, it is clear that the right-hand side of (3.3) is continuous with respect to $\alpha$. The proof is completed. \hfill \square

Remark 3.3. If $E$ possesses a higher regularity, i.e., $E \in C^1(H, \mathbb{R})$ and $E' : H \rightarrow H^*$ is locally $\gamma_1$-Hölder continuous for some $\gamma_1 \in (0, 1)$, then the regularity assumption of $p$ in Lemma 3.2 can be relaxed to local $\gamma_2$-Hölder continuity around $v(\alpha)$ for some $\gamma_2 \in (1/(\gamma_1 + 1), 1]$. Actually, the key step in the proof is to justifying that the right-hand side of (3.4) is $o(s||d||)$. The local Hölder continuity of $E'$ and $p$ implies that there exist two constants $C_1(\alpha), C_2(\alpha) > 0$ such that when $s \rightarrow 0$,

$$
\|E'(\xi) - E'(p(\alpha)))\|_H, \|p(v(\alpha + s)) - p(v(\alpha))||
$$
Definition 3.4 (Descent direction). Let \( p \) be a local peak selection with respect to \( L \) at \( v \in S_H \setminus L \). Assume that \( E \in C^1(H, \mathbb{R}) \) and \( E'(p(v)) \neq 0 \). A vector \( d \in [L, v]^+ \) is called a descent direction of \( E \) with respect to \( L \) at \( p(v) \) if \( \langle E'(p(v)), d \rangle < 0 \).

Remark 3.5. It is noted that the above definition is slightly different from the usual definition of the descent direction in the optimization theory. In fact, the descent direction \( d \) in Definition 3.4 is required not only to decrease the functional \( E \) at \( p(v) \) (i.e., \( \langle E'(p(v)), d \rangle < 0 \)), but also to satisfy the orthogonality condition \( d \perp [L, v] \). We make some comments on the reasons for introducing the latter condition as follows:

(i) The descent direction \( d \) in Definition 3.4 is for the outer-level minimization and should be relatively independent of the inner-level maximization. Intuitively, an iteration along a descent direction in \([L, v]^+\), i.e., a descent direction with the component zero in \([L, v]\), tends to enhance the stability of the algorithm.

(ii) By Lemma 2.2, the condition \( d \in [L, v]^+ \) is automatically satisfied for the steepest descent direction \( d^{SD} = -\nabla E(p(v)) \) at \( p(v) \), i.e., the Riesz representer of \(-E'(p(v))\) determined by \( (d^{SD}, \phi) = -\langle E'(p(v)), \phi \rangle, \forall \phi \in H \). Actually, this orthogonality plays a very important role in both the algorithm implementation and the theoretical analysis of the classical LMMs. Based on this observation, preserving this orthogonality to the general descent direction \( d \) for the outer-level minimization is a preferred choice.

Definition 3.6 (Normalized Wolfe-Powell-type step-size search rules). Suppose \( E \in C^1(H, \mathbb{R}) \) and let \( v \in S_H \setminus L \), \( p \) be a peak selection of \( E \) with respect to \( L \), and \( d \in [L, v]^+ \) be a descent direction of \( E \) with respect to \( L \) at \( p(v) \). Define \( p(u) = t_u w + w_L^+ \) \((u \in S_H)\) with \( t_u \geq 0 \) and \( w_L^+ \in L \). For two given constants \( \sigma_1 \) and \( \sigma_2 \) with \( 0 < \sigma_1 < \sigma_2 < 1 \), we say that the step-size \( \alpha > 0 \) satisfies

- the normalized Wolfe-Powell-type step-size search rule at \( v \), if it holds that
  \[
  E(p(v(\alpha))) \leq E(p(v)) + \sigma_1 \alpha t_v \langle E'(p(v)), d \rangle, \tag{3.5a}
  \]
  \[
  \hat{t}_v(\alpha) \left| \langle E'(p(v(\alpha))), d \rangle \right| \geq \sigma_2 t_v \langle E'(p(v)), d \rangle, \tag{3.5b}
  \]
  where \( \hat{t}_v(\alpha) = t_v(\alpha)/\sqrt{1 + \alpha^2 \|d\|^2} \);

- the normalized strong Wolfe-Powell-type step-size search rule at \( v \), if it holds that
  \[
  E(p(v(\alpha))) \leq E(p(v)) + \sigma_1 \alpha t_v \langle E'(p(v)), d \rangle, \tag{3.6a}
  \]
  \[
  \hat{t}_v(\alpha) \left| \langle E'(p(v(\alpha))), d \rangle \right| \leq -\sigma_2 t_v \langle E'(p(v)), d \rangle. \tag{3.6b}
  \]

The condition (3.5a) or (3.6a) is referred to as the sufficient decrease condition, while the conditions (3.5b) and (3.6b) are referred to as curvature conditions. It is pointed out that if the steepest descent direction is employed, (3.5a) or (3.6a) is equivalent to the normalized Armijo-type condition used in traditional LMMs [18,32,33].

The feasibility of normalized (strong) Wolfe-Powell-type step-size search rules above is provided as follows.
Theorem 3.7. Let $E \in C^1(H, \mathbb{R})$, $v \in S_H$, $d \in [L, v]^\perp$ and $p$ be a peak selection of $E$ with respect to $L$. Define $p(u) = t_u u + w_u^L$, $u \in S_H$ with $t_u \geq 0$ and $w_u^L \in L$. Assume that (i) $p$ is locally Lipschitz continuous on the curve $\{v(\alpha) : \alpha \geq 0\}$, (ii) $t_0 > 0$, (iii) $d$ is a descent direction of $E$ with respect to $L$ at $p(v)$, i.e., $\langle E(p(v)), d \rangle < 0$, and (iv) $\inf_{\alpha > 0} E(p(v(\alpha))) > -\infty$. Then for given $\sigma_1$ and $\sigma_2$ with $0 < \sigma_1 < \sigma_2 < 1$, there exist two positive constants $\alpha_1$ and $\alpha_2$ with $\alpha_1 < \alpha_2$ such that for any $\alpha \in (\alpha_1, \alpha_2)$, it satisfies the normalized strong Wolfe-Powell-type step-size search rule (3.6) and therefore the normalized Wolfe-Powell-type step-size search rule (3.5).

Proof. Since (3.6) yields (3.5), we only need to verify that there exists an interval $(\alpha_1, \alpha_2)$ such that (3.6) holds for all $\alpha \in (\alpha_1, \alpha_2)$. Set $\varphi(\alpha) := E(p(v(\alpha)))$, $\alpha \geq 0$ and $\varphi(0) = E(p(v))$. In view of Lemmas 2.3 and 3.2, we have $\varphi \in C^1([0, \infty), \mathbb{R})$ and $\varphi'(\alpha) = \hat{t}_{v(\alpha)} \langle E'(p(v(\alpha))), d \rangle$ with $\hat{t}_{v(\alpha)} = t_{v(\alpha)}/\sqrt{1 + \alpha^2 \|d\|^2}$. Note that the conditions (ii) and (iii) imply $\varphi'(0) = t_v(E'(p(v)), d) < 0$. Hence, (3.6) can be rewritten as

$$
\varphi(\alpha) \leq \varphi(0) + \sigma_1 \alpha \varphi'(0), \quad |\varphi'(\alpha)| \leq -\sigma_2 \varphi'(0), \quad \alpha \geq 0,
$$

and for all $\alpha > 0$ small enough, $\varphi(\alpha) < \varphi(0) + \sigma_1 \alpha \varphi'(0)$ holds. In addition, since $\varphi(0) + \sigma_1 \alpha \varphi'(0) \to -\infty$ as $\alpha \to +\infty$ and the condition (iv) states that $\varphi(\alpha)$ is bounded from below for all $\alpha > 0$, apparently $\varphi(\alpha) > \varphi(0) + \sigma_1 \alpha \varphi'(0)$ holds for all $\alpha > 0$ large enough. Consequently, the equation

$$
\varphi(\alpha) = \varphi(0) + \sigma_1 \alpha \varphi'(0), \quad \alpha > 0
$$

admits at least one positive solution. Let $\bar{\alpha} > 0$ be the smallest positive solution of the equation (3.8). Then it implies that

$$
\varphi(\alpha) < \varphi(0) + \sigma_1 \alpha \varphi'(0), \quad \forall \alpha \in (0, \bar{\alpha}).
$$

From the mean value theorem, there exists an $\bar{\alpha} \in (0, \bar{\alpha})$ such that $\varphi(\bar{\alpha}) - \varphi(0) = \varphi'(\bar{\alpha}) \delta$, which leads to $\varphi'(\bar{\alpha}) = \sigma_1 \varphi'(0)$ by the definition of $\bar{\alpha}$. The facts that $0 < \sigma_1 < \sigma_2 < 1$ and $\varphi'(0) < 0$ imply that $\sigma_2 \varphi'(0) < \sigma_1 \varphi'(0) = \varphi'(\bar{\alpha}) < 0$. Therefore, $|\varphi'(\bar{\alpha})| < -\sigma_2 \varphi'(0)$. By the continuity of $\varphi'(\alpha)$, there exists $\delta \in (0, \min\{\bar{\alpha}, \bar{\alpha} - \delta\})$ such that

$$
|\varphi'(\alpha)| < -\sigma_2 \varphi'(0), \quad \forall \alpha \in (\bar{\alpha} - \delta, \bar{\alpha} + \delta) \subset (0, \bar{\alpha}).
$$

Setting $\alpha_1 = \bar{\alpha} - \delta$ and $\alpha_2 = \bar{\alpha} + \delta$, we see that the combination of (3.9) and (3.10) states that (3.7) holds for all $\alpha \in (\alpha_1, \alpha_2)$. □

Remark 3.8. Figure 3 provides a geometric interpretation of the feasibility of the normalized (strong) Wolfe-Powell-type step-size search rules.
3.2 Normalized Wolfe-Powell-type local minimax algorithm

Following the idea of traditional Wolfe-Powell LMMs [17,18,32], we describe the vital steps of the NWP-LMM in Algorithm 1.

Algorithm 1 Normalized Wolfe-Powell-type local minimax algorithm

Step 1. Take constants $\varepsilon > 0$, $\sigma_1$ and $\sigma_2$ with $0 < \sigma_1 < \sigma_2 < 1$, and $n - 1$ previously found critical points $u_1, u_2, \ldots, u_{n-1}$ of $E$, where $u_{n-1}$ is the one with the highest critical value in $\{u_k\}$ $(1 \leq k \leq n - 1)$. Set $L = \text{span}\{u_1, u_2, \ldots, u_{n-1}\}$, let $k := 0$, and choose an initial ascent direction $v_0 = v^+_0 + v^-_0 \in S_H$ at $u_{n-1}$ with $v^+_0 \in L$, $v^-_0 \in L^\perp$ and $v^-_0 \neq 0$. With an initial guess $w = v_0 + u_{n-1}$, solve

$$w_0 = \arg \max_{w \in [L,v_0]} E(u),$$

and define $w_0 = p(v_0) = t_0 v_0 + w^L_0$, where $t_0 \geq 0$ and $w^L_0 \in L$.

Step 2. Compute a descent direction $d_k \in [L,v_k]^\perp$ of $E$ with respect to $L$ at $w_k = p(v_k)$ such that $\langle E'(w_k), d_k \rangle < 0$, which will be discussed in Section 5.

Step 3. If the stopping criterion $\|E'(w_k)\|_{H^*} < \varepsilon$ is satisfied (or more criteria are satisfied if necessary), then output $u_n = w_k$ and stop; otherwise, go to Step 4.

Step 4. Set $v_k(\alpha) = \frac{\alpha v_k + d_k}{\|\alpha v_k + d_k\|}$ and find a step-size $\alpha_k > 0$ satisfying the normalized Wolfe-Powell-type step-size search rule, i.e.,

$$E(p(v_k(\alpha_k))) \leq E(w_k) + \sigma_1 \alpha_k t_k \langle E'(w_k), d_k \rangle, \quad \text{(3.11a)}$$

$$t_k \langle E'(p(v_k(\alpha_k))), d_k \rangle \geq \sigma_2 t_k \langle E'(w_k), d_k \rangle, \quad \text{(3.11b)}$$

or the normalized strong Wolfe-Powell-type step-size search rule, i.e.,

$$E(p(v_k(\alpha_k))) \leq E(w_k) + \sigma_1 t_k \alpha_k \langle E'(w_k), d_k \rangle, \quad \text{(3.12a)}$$

$$t_k \langle E'(p(v_k(\alpha_k))), d_k \rangle \leq -\sigma_2 t_k \langle E'(w_k), d_k \rangle, \quad \text{(3.12b)}$$

where $p(v_k(\alpha_k)) = t_k(\alpha_k) v_k(\alpha_k) + w^L_k(\alpha_k)$ with $t_k(\alpha_k) \geq 0$ and $w^L_k(\alpha_k) \in L$ is the local maximizer of $E$ on $[L,v_k(\alpha_k)]$ computed by utilizing $w = t_k(\alpha_k) v_k(\alpha_k) + w^L_k(\alpha_k)$ as an initial guess and $t_k(\alpha_k) = t_k(\alpha_k) / \sqrt{1 + \alpha_k^2 \|d_k\|^2}$.

Step 5. Set $v_{k+1} = v_k(\alpha_k)$, $t_{k+1} = t_k(\alpha_k)$, $w^L_{k+1} = w^L_k(\alpha_k)$ and $w_{k+1} = p(v_{k+1}) = t_{k+1} v_{k+1} + w^L_{k+1}$. Then update $k := k + 1$ and go to Step 2.

We remark here that similar to the classical (strong) Wolfe-Powell line search algorithm in the optimization theory (see, e.g., [12,23,27]), one can employ an interpolation approach to efficiently find a step-size $\alpha_k > 0$ satisfying the normalized Wolfe-Powell-type step-size search rule (3.11) or the normalized strong Wolfe-Powell-type step-size search rule (3.12) in Step 4 of Algorithm 1. The implementation details are skipped here for brevity.

4 Global convergence

In this section, we establish the global convergence of Algorithm 1. In order to achieve this goal, the following concept of compactness is needed. We simply utilize the same notations as those in Algorithm 1 throughout this section.

Definition 4.1 (See [25]). A functional $E \in C^1(H, \mathbb{R})$ is said to satisfy the Palais-Smale (PS) condition if every sequence $\{w_n\} \subset H$ such that $\{E(w_n)\}$ is bounded and $E'(w_n) \to 0$ in $H^*$ has a convergent subsequence.

It is pointed out that the following lemma gives a significant behavior of the sequence generated by Algorithm 1, which does not depend on the choice of descent directions and step-sizes in the algorithm. The proof is similar to that of [32, Lemma 2.3] and is omitted here for brevity.

Lemma 4.2. Let $\{v_k\}$ be a sequence generated by Algorithm 1 with $v_0 \in S\setminus L$. Define $v_k = v^+_k + v^-_k$ with $v^+_k \in L^\perp$ and $v^-_k \in L$, $k = 0, 1, \ldots$. Then it holds that $\|v^+_k\| \leq \|v^-_k\| \leq 1$ and $v^+_k = \tau_k v^-_k$ with $0 < \tau_{k+1} \leq \tau_k \leq 1$ for $k = 0, 1, \ldots$.

Lemma 4.2 states that once an initial ascent direction $v_0 = v^+_0 + v^-_0 \in S\setminus L$ is chosen in Algorithm 1, the closed subset

$$\mathcal{V}_0 := \{v = v^+_0 + \tau v^-_0 \in S_H : v^+_0 \in L^\perp, 0 \leq \tau \leq 1\} \subset S_{[L^\perp, v^+_0]} \subset S_H \setminus L \quad \text{(4.1)}$$
contains all the possible vectors \( v_k \) that Algorithm 1 may generate. Thus, the domain of a peak selection can be limited to \( V_0 \) instead of \( S_H \).

We remark here that in general, the local peak selection \( p(v) \) defined on \( V_0 \) is no longer a homeomorphism. The following weak version related to the homeomorphism property of the local peak selection \( p(v) \) plays a significant role in establishing the global convergence. It can be verified by an analogous argument to that of [32, Theorem 2.1] and only the continuity of the peak selection \( p(v) \) on \( V_0 \) is sufficient. Consequently, we skip the proof for simplicity.

**Lemma 4.3.** Suppose \( E \in C^1(H, \mathbb{R}) \) and let \( p \) be a peak selection of \( E \) with respect to \( L \) satisfying (i) \( p \) is continuous on \( V_0 \) and (ii) \( t_k \geq \delta \) for some \( \delta > 0, \forall k = 0, 1, \ldots \). If the sequence \( \{ w_k \} \) generated by Algorithm 1 contains a subsequence \( \{ w_{k_i} \} \) converging to some \( u_* \in H \), then the corresponding subsequence \( \{ v_{k_i} \} \) converges to some \( v_* \in V_0 \) with \( u_* = p(v_*) \).

Before proving the global convergence of Algorithm 1, we give some assumptions on the general descent direction, which are quite reasonable and hold for many descent directions, especially for the steepest descent direction used in traditional LMMs.

(A1) \( \langle E'(w_k), d_k \rangle \leq -c_1 \| E'(w_k) \|^2_H \), for some \( c_1 > 0, \forall k = 0, 1, \ldots \)

(A2) \( \| d_k \| \leq c_2 \| E'(w_k) \|^2_H \), for some \( c_2 > 0, \forall k = 0, 1, \ldots \)

(A3) If \( \{ v_k \} \) contains a subsequence \( \{ v_{k_i} \} \) converging to some \( \bar{v} \) with \( E'(p(\bar{v})) \neq 0 \), then the corresponding descent direction subsequence \( \{ d_{k_i} \} \) converges.

Here, the assumption (A1) serves as a strong descent condition, and the assumption (A2) requires the decreasing condition (3.11a) (or (3.12a)), the condition (ii) and the assumption (A1) say that

\[
\sum_{k=0}^{\infty} \alpha_k \| E'(w_k) \|^2_H < \infty.
\]

Furthermore, if the functional \( E \) satisfies the (PS) condition, then

(d) \( \{ w_k \} \) contains a subsequence converging to a critical point \( u_* \notin L \). In addition, if \( u_* \) is isolated, then \( w_k \to u_* \) as \( k \to \infty \).

**Proof.** The decreasing condition (3.11a) (or (3.12a)), the condition (ii) and the assumption (A1) say that for \( k = 0, 1, \ldots \),

\[
E(w_{k+1}) - E(w_k) \leq \sigma_1 \alpha_k t_k \langle E'(w_k), d_k \rangle \leq -\sigma_1 \delta c_1 \alpha_k \| E'(w_k) \|^2_H. \tag{4.2}
\]

Therefore, the sequence \( \{ E(w_k) \} \) is monotonically non-increasing. In addition, since the condition (iii) guarantees that \( E(w_k) \) is bounded from below for all \( k = 0, 1, \ldots \), the sequence \( \{ E(w_k) \} \) converges to some \( E_\infty := \inf_{k \geq 0} E(w_k) \).

Then adding up (4.2), we can arrive at

\[
\sum_{k=0}^{\infty} (E(w_{k+1}) - E(w_k)) \leq -\sigma_1 \delta c_1 \sum_{k=0}^{\infty} \alpha_k \| E'(w_k) \|^2_H. \tag{4.3}
\]

Hence, the left-hand side of (4.3) converges to \( E_\infty - E(w_0) \) which is finite. This immediately leads to the conclusion (a).

To prove the conclusion (b), let \( \bar{u} \in H \) be an accumulation point of the sequence \( \{ w_k \} \). Then there exists a subsequence \( \{ w_{k_i} \} \) converging to \( \bar{u} \) as \( i \to \infty \). Recalling the weak version of the homeomorphism property in Lemma 4.3, we see that the corresponding subsequence \( \{ v_{k_i} \} \) converges to some \( \bar{v} \in V_0 \) satisfying \( \bar{u} = p(\bar{v}) = t_0 \bar{v} + w_0^v \). Lemma 4.2 and the condition (ii) yield that for some \( \delta > 0 \),

\[
\text{dist}(\bar{u}, L) = \lim_{i \to \infty} \text{dist}(w_{k_i}, L) = \lim_{i \to \infty} t_k \| v_{k_i}^v \| \geq \delta \| v_0^v \| > 0.
\]
and therefore $\bar{u} \notin L$.

The following is to verify that $\bar{u}$ is a critical point by taking full advantage of the curvature condition (3.11b) (or (3.12b)), which states that

$$
\frac{t_{k+1}}{\sqrt{1 + \alpha_k^2 \|d_k\|^2}} (E'(w_{k+1}), d_k) \geq \sigma_2 t_{k+1} (E'(w_k), d_k), \quad i = 0, 1, \ldots
$$

(4.4)

By the contradiction argument, suppose that $\bar{u}$ is not a critical point, i.e., $E'(\bar{u}) \neq 0$. Since the functional $E \in C^1(H, \mathbb{R})$, one can obtain

$$
E'(w_k) \to E'(\bar{u}) \quad \text{as} \quad i \to \infty.
$$

(4.5)

As a result, for all $i$ large enough, $\|E'(w_k)\|_{H^*} > \|E'(\bar{u})\|_{H^*/2} > 0$. In view of the conclusion (a), it leads to

$$
\alpha_k \to 0 \quad \text{as} \quad i \to \infty.
$$

(4.6)

In addition, the assumption (A3) states that there exists a $\bar{d} \in [L, \bar{v}]^\perp$ such that

$$
d_k \to \bar{d} \quad \text{as} \quad i \to \infty.
$$

(4.7)

Therefore, (4.6) and (4.7) immediately indicate that

$$
v_{k+1} = \frac{v_k + \alpha_k d_k}{\sqrt{1 + \alpha_k^2 \|d_k\|^2}} \to \bar{v} \quad \text{in} \quad H \quad \text{as} \quad i \to \infty
$$

and

$$
E'(w_{k+1}) = E'(p(v_{k+1})) \to E'(p(\bar{v})) = E'(\bar{u}) \quad \text{as} \quad i \to \infty.
$$

(4.8)

by the continuity of $E'$ and $p$. Moreover, reviewing Lemma 2.3 and the condition (ii), for some $\delta > 0$, we have

$$
t_k \to t_\bar{v} \geq \delta > 0 \quad \text{and} \quad t_{k+1} \to t_\bar{v} \geq \delta > 0 \quad \text{as} \quad i \to \infty.
$$

(4.9)

Above all, combining (4.5)–(4.9) and taking $i \to \infty$ in (4.4) imply

$$
t_\bar{v} \langle E'(\bar{u}), \bar{d} \rangle \geq \sigma_2 t_\bar{v} \langle E'(\bar{u}), \bar{d} \rangle.
$$

(4.10)

Since $E'(\bar{u}) = E'(p(\bar{v})) \neq 0$, it follows from the assumption (A1), (4.5) and (4.7) that $\langle E'(\bar{u}), \bar{d} \rangle \leq -c_1 \|E'(\bar{u})\|_{H^*} < 0$ for some $c_1 > 0$. Thus, we have $t_\bar{v} \langle E'(\bar{u}), \bar{d} \rangle < 0$ and (4.10) contradicts the fact that $0 < \sigma_2 < 1$. Consequently, the accumulation point $\bar{u} = p(\bar{v})$ is a critical point. The conclusion (b) is obtained.

Next, we prove the conclusion (c) by the contradiction argument. Suppose that

$$
\delta_1 := \liminf_{k \to \infty} \|E'(w_k)\|_{H^*} > 0.
$$

Then for $k$ large enough, $\|E'(w_k)\|_{H^*} \geq \delta_1/2 > 0$. Thus, the conclusion (a) admits

$$
\sum_{k=0}^{\infty} \alpha_k \|E'(w_k)\|_{H^*} < \infty.
$$

(4.11)

In addition, Lemma 3.1 and the assumption (A2) yield that for some $c_2 > 0$,

$$
\|u_{k+1} - u_k\| = \|v_k (\alpha_k) - v_k\| \leq \alpha_k \|d_k\| \leq c_2 \alpha_k \|E'(w_k)\|_{H^*}, \quad k = 0, 1, \ldots
$$

(4.12)

Combining (4.11) and (4.12) results in $\sum_{k=0}^{\infty} \|u_{k+1} - u_k\| < \infty$. Therefore, $\{v_k\}$ is a Cauchy sequence in the closed subset $\bar{V}_0$. Immediately, the completeness of the closed subset $\bar{V}_0$ implies that there exists a $\bar{v} \in \bar{V}_0$ such that $v_k \to \bar{v}$ as $k \to \infty$. Furthermore, by the continuity of $p$ and $E'$, we have $w_k = p(v_k) \to p(\bar{v})$ with $p(\bar{v})$ the accumulation point, and $E'(w_k) \to E'(p(\bar{v}))$ as $k \to \infty$. Hence, it holds that

$$
\|E'(p(\bar{v}))\|_{H^*} = \lim_{k \to \infty} \|E'(w_k)\|_{H^*} = \delta_1 > 0.
$$
This is a contradiction to the conclusion (b). Thus, the conclusion (c) is proved.

The rest is to prove the conclusion (d). Since \(\{E(w_k)\}\) converges to \(E_\infty\) by the proof of the conclusion (a), according to the conclusion (c), there exists a subsequence \(\{w_{k_i}\}\) such that \(E(w_{k_i}) \to E_\infty\) and \(E'(w_{k_i}) \to 0\) as \(i \to \infty\). By the (PS) condition, \(\{w_{k_i}\}\) possesses a subsequence, still denoted by \(\{w_{k_i}\}\), which converges to a critical point \(u_* \in H\). In addition, according to the conclusion (b), \(u_* \notin L\) holds. Finally, under the assumption that \(u_*\) is isolated, following the analogous lines in the proof of [39, Theorem 2.4] for the global convergence, we complete the proof. □

5 Descent directions

In this section, we propose two specific types of descent directions for implementing Algorithm 1 in detail. One is the PSD direction, and the other is the CG-type descent direction. We use the same notations as those in Algorithm 1 for subsequent discussions in this section, unless otherwise specified.

5.1 Preconditioned steepest descent direction

The gradient of \(E\) at \(w_k = p(v_k)\), denoted by \(g_k = \nabla E(w_k)\), is defined by

\[
(g_k, \phi) = (E'(p(v_k)), \phi), \quad \forall \phi \in H.
\]

From the Riesz representation theorem, the gradient \(g_k \in H\) exists uniquely and \(\|g_k\| = \|E'(p(v_k))\|_{H^*}\).

Consider the following PSD direction:

\[
d_k = -T_k g_k, \quad k = 0, 1, \ldots, \tag{5.2}
\]

where \(T_k = T(v_k)\), called a preconditioner at \(v_k\), is a positive-definite and self-adjoint bounded linear operator on \(H\) with an invariant subspace \([L, v_k]^\perp\). Assume that \(T_k = T(v_k)\) satisfies

- (T1) for some \(c_3 > 0\), \(\|T(v_k)u\| \leq c_3 \|u\|, \forall u \in [L, v_k]^\perp\);
- (T2) for some \(c_4 > 0\), \(\langle T(v_k)u, u \rangle \geq c_4 \|u\|^2, \forall u \in [L, v_k]^\perp\);
- (T3) \(T(v)\) is continuous at the accumulation point \(\bar{v}\) of \(\{v_k\}\) such that \(E'(p(\bar{v})) \neq 0\).

The following theorem provides the global convergence result of Algorithm 1 with \(d_k\) taken as the PSD direction (5.2).

**Theorem 5.1.** Suppose \(E \in C^4(H, \mathbb{R})\) and let \(p\) be a peak selection of \(E\) with respect to \(L\), and \(\{v_k\}\) and \(\{w_k\}\) be sequences generated by Algorithm 1 with \(d_k = -T_k g_k\) the PSD direction defined in (5.2) and \(T_k = T(v_k)\) satisfying the assumptions (T1)–(T3). If (i) \(p\) is locally Lipschitz continuous on \(V_0\), (ii) \(t_k \geq \delta\) for some \(\delta > 0\), \(\forall k = 0, 1, \ldots\), and (iii) \(\inf_{v \in V_0} E(p(v)) > -\infty\), then those conclusions (a)–(d) in Theorem 4.4 hold.

Proof. Since \(g_k \in [L, v_k]^\perp\) from Lemma 2.2 and \([L, v_k]^\perp\) is an invariant subspace of \(T_k = T(v_k)\), we have \(d_k = -T_k g_k \in [L, v_k]^\perp\). In order to prove the conclusion, it suffices to verify the assumptions (A1)–(A3). In fact, firstly, the assumption (T2) states that

\[
\langle E'(w_k), d_k \rangle = -(g_k, T_k g_k) \leq -c_4 \|g_k\|^2 = -c_4 \|E'(w_k)\|^2_{H^*}, \quad k = 0, 1, \ldots, \tag{5.3}
\]

which is (A1) (with \(c_4 = c_1\)). Furthermore, by the assumption (T1),

\[
\|d_k\| = \|T_k g_k\| \leq c_3 \|g_k\| = c_3 \|E'(w_k)\|_{H^*}, \quad k = 0, 1, \ldots \tag{5.4}
\]

Thus, (A2) is verified by taking \(c_3 = c_2\). Finally, (A3) directly follows from the assumption (T3) and the continuity of \(E'\) and \(p\). □

Taking \(T_k\) in Theorem 5.1 simply as the identity operator on \(H\) yields the following corollary, which draws the global convergence of Algorithm 1 with the standard steepest descent direction utilized at each iterative step.
Corollary 5.2. Suppose $E \in C^1(H, \mathbb{R})$ and let $p$ be a peak selection of $E$ with respect to $L$, and $\{v_k\}$ and $\{w_k\}$ be sequences generated by Algorithm 1 with $d_k = -g_k$. If (i) $p$ is locally Lipschitz continuous on $V_0$, (ii) $t_k \geq \delta$ for some $\delta > 0$, $\forall k = 0, 1, \ldots$, and (iii) $\inf_{v \in V_0} E(p(v)) > -\infty$, then those conclusions (a)–(d) in Theorem 4.4 hold.

Remark 5.3. It is noted that [39, Theorem 2.4], [20, Theorem 5.1] and Corollary 5.2 provide, respectively, the global convergence of the NA-LMM, NG-LMM and NWP-LMM with the steepest descent direction. Consequently, mathematical justifications for LMMs combined with several typical inexact normalized step-size search rules for the steepest descent direction have been systematically established.

5.2 Conjugate-gradient-type direction

For $k = 0, 1, \ldots$, define $v_k = v^L_k + v^w_k$ with $v^L_k \in L$ and $0 \neq v^w_k \in L^\perp$. Similar to the construction of the nonlinear CG method in the optimization theory (see, e.g., [7, 14]), we consider the following CG-type direction for Algorithm 1:

$$d_k = \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_k E_k d_{k-1}, & k \geq 1. \end{cases} \quad (5.5)$$

Here, $\beta_k \in \mathbb{R}$ is a parameter to be determined, $g_k = \nabla E(w_k)$ is the gradient of $E$ at $w_k$ defined in (5.1), and $E_k$ is the orthogonal projection from $H$ onto $[L, v_k]^\perp$. Since $g_k \in [L, v_k]^\perp$ from Lemma 2.2, we have $d_k \in [L, v_k]^\perp$ for all $k \geq 0$. In addition, according to the definition of $E_k$, it holds that $(g_k, E_k d_{k-1}) = (g_k, d_{k-1})$, $k \geq 1$. Thus, the CG-type direction (5.5) satisfies

$$(g_k, d_k) = -\|g_k\|^2 + \beta_k (g_k, d_{k-1}), \quad k \geq 1. \quad (5.6)$$

Remark 5.4. We remark here that $E_k d_{k-1}$ can be explicitly expressed as

$$E_k d_{k-1} = d_{k-1} - \|v^w_k\|^2 (d_{k-1}, v^w_k) v^w_k, \quad k \geq 1. \quad (5.7)$$

Actually, by applying the facts that $d_{k-1} \in [L, v_{k-1}]^\perp \subset L^\perp$, $E_k d_{k-1} \in [L, v_k]^\perp \subset L^\perp$, and $\text{Id} - E_k$ is an orthogonal projection onto $([L, v_k]^\perp)^\perp = L \oplus [v^w_k]$, where $\text{Id}$ is the identity operator on $H$ and $\oplus$ denotes the direct sum, we can conclude that

$$d_{k-1} - E_k d_{k-1} \in ((L \oplus [v^w_k]) \cap L^\perp) = [v^w_k].$$

Thus, $d_{k-1} - E_k d_{k-1} = c v^w_k$ for some $c \in \mathbb{R}$. Taking the inner product with $v^w_k$ and noting that $(E_k d_{k-1}, v^w_k) = (E_k d_{k-1}, v^w_k - v^w_k) = 0$, we can obtain

$$(d_{k-1}, v^w_k) = (d_{k-1} - E_k d_{k-1}, v^w_k) = c \|v^w_k\|^2,$$

yielding $c = \|v^w_k\|^2 (d_{k-1}, v^w_k)$. Consequently, the expression (5.7) is true.

The following lemma shows that if the exact step-size search rule is applied in the previous iteration, the CG-type direction (5.5) with an arbitrary parameter $\beta_k$ is a descent direction.

Lemma 5.5. For $k \geq 1$, let $v_k = v_{k-1}(\alpha_{k-1})$ with $d_{k}$ defined in (5.5) and $\alpha_{k-1} > 0$ be a local minimizer of $E(p(v_{k-1}(\alpha)))$ along $\{\alpha : \alpha > 0\}$. If $E \in C^1(H, \mathbb{R})$, $p$ is locally Lipschitz continuous around $v_k$ and $t_k > 0$, then $(g_k, d_k) = -\|g_k\|^2$.

Proof. From Lemma 3.2, $E(p(v_{k-1}(\alpha)))$ is continuously differentiable at $\alpha = \alpha_{k-1}$ and

$$\left. \frac{d}{d\alpha} E(p(v_{k-1}(\alpha))) \right|_{\alpha = \alpha_{k-1}} = \frac{t_k}{\sqrt{1 + \alpha_{k-1}^2 \|d_{k-1}\|^2}} (E'(p(v_k)), d_{k-1}).$$

By applying the facts that $t_k > 0$ and $E(p(v_{k-1}(\alpha)))$ attains its local minimum at $\alpha_{k-1}$, we have $E'(p(v_k)), d_{k-1}) = (g_k, d_{k-1}) = 0$. Thus, the conclusion follows from (5.6) immediately. \qed
Due to the fact that the exact step-size search rule is quite expensive in practical computations, we focus on ensuring that the CG-type direction $d_k$ defined in (5.5) is a descent direction when a suitable inexact step-size search rule is used.

Inspired by the well-known Fletcher-Reeves CG method [13] in the optimization theory, we set

$$
\beta_k^{\text{FR-like}} = \frac{\gamma_k \|g_k\|^2}{\|g_{k-1}\|^2}, \quad k \geq 1,
$$

(5.8)

where $\gamma_k = \hat{t}_k/t_{k-1}$ with $t_k = t_k/\sqrt{1 + \alpha_{k-1}^2 \|d_{k-1}\|^2}$. It can be verified that the CG-type direction (5.5) with $\beta_k = \beta_k^{\text{FR-like}}$ is a descent direction if the step-size $\alpha_k$ in each iteration satisfies the normalized strong Wolfe-Powell-type step-size search rule (3.12) with $\sigma_2 \in (0, 1/2)$, as stated in the following lemma. The proof follows the lines of the proof of [1, Theorem 1].

**Lemma 5.6.** For $k = 0, 1, \ldots$, if $g_k \neq 0$, $d_k$ in Algorithm 1 is defined in (5.5) with $\beta_k = \beta_k^{\text{FR-like}}$, and $\alpha_k$ is determined by the normalized strong Wolfe-Powell-type step-size search rule (3.12) with $\sigma_2 \in (0, 1/2)$, then

$$
-\frac{1 - \sigma_2^{k+1}}{1 - \sigma_2} \leq \frac{\langle g_k, d_k \rangle}{\|g_k\|^2} \leq -1 + \frac{1 - 2\sigma_2 + \sigma_2^{k+1}}{1 - \sigma_2}, \quad k = 0, 1, \ldots
$$

(5.9)

Consequently,

$$
\langle g_k, d_k \rangle < 0, \quad k = 0, 1, \ldots
$$

(5.10)

**Proof.** For $k = 0$ and $d_0 = -g_0$, the conclusions (5.9) and (5.10) are immediate. According to the inductive argument, suppose that the conclusions (5.9) and (5.10) hold for $k - 1$, $k \geq 1$. In view of the definition of $\beta_k^{\text{FR-like}}$ in (5.8), the fact (5.6) yields

$$
\frac{\langle g_k, d_k \rangle}{\|g_k\|^2} = -1 + \gamma_k \frac{\langle g_k, d_{k-1} \rangle}{\|g_{k-1}\|^2}, \quad k \geq 1.
$$

Using the inductive assumption (5.10) for $k - 1$, $k \geq 1$, the normalized strong Wolfe-Powell-type step-size search rule (3.12) states that $\gamma_k \|g_k, d_{k-1}\| \leq -\sigma_2(g_{k-1}, d_{k-1})$, and therefore,

$$
-1 + \sigma_2 \frac{\langle g_{k-1}, d_{k-1} \rangle}{\|g_{k-1}\|^2} \leq \frac{\langle g_k, d_k \rangle}{\|g_k\|^2} \leq -1 - \sigma_2 \frac{\langle g_{k-1}, d_{k-1} \rangle}{\|g_{k-1}\|^2}.
$$

Furthermore, by the inductive assumption (5.9) for $k - 1$, $k \geq 1$, we can arrive at

$$
-\frac{1 - \sigma_2^{k+1}}{1 - \sigma_2} = -1 - \sigma_2 \frac{1 - \sigma_2}{1 - \sigma_2} \leq \frac{\langle g_k, d_k \rangle}{\|g_k\|^2} \leq -1 + \sigma_2 \frac{1 - \sigma_2}{1 - \sigma_2} = -1 + \frac{1 - 2\sigma_2 + \sigma_2^{k+1}}{1 - \sigma_2},
$$

which leads to the conclusion (5.9) for $k \geq 1$. Since $\sigma_2 \in (0, 1/2)$, it immediately follows that $\langle g_k, d_k \rangle < 0$, $k \geq 1$. The proof is finished by the inductive argument. \hfill \Box

**Remark 5.7.** Under the assumptions in Lemma 5.6, we have from (5.9) that $\langle g_k, d_k \rangle \leq -c_1 \|g_k\|^2$ with $c_1 = (1 - 2\sigma_2)/(1 - \sigma_2) > 0$, i.e., the CG-type direction (5.5) with $\beta_k = \beta_k^{\text{FR-like}}$ satisfies the assumption (A1). Moreover, under the same assumptions, these conclusions in Lemma 5.6 can be extended directly to any choice of $\beta_k$ satisfying $|\beta_k| \leq \beta_k^{\text{FR-like}}$.

**Remark 5.8.** It is currently unclear whether the assumptions (A2) and (A3) hold for the NWP-LMM with the CG-type descent direction (5.5). Thus, the global convergence of it has not been verified yet and will be our future work. Indeed, the NWP-LMM with the CG-type descent direction (5.5) is very efficient for finding multiple solutions of semilinear elliptic PDEs, compared with the traditional LMMs and the NWP-LMM with the steepest descent direction, which will be shown in Section 6. Actually, several different constructions of CG-type descent directions can also be designed based on the similar ideas of various CG methods in the optimization theory, which can be found, e.g., in [7, 14].
6 Numerical examples

In this section, we apply our NWP-LMM to find multiple unstable solutions of the semilinear elliptic boundary value problem (BVP)

\[ \begin{aligned}
-\Delta u(x) + a(x)u(x) &= f(x, u(x)), & x &\in \Omega, \\
 u(x) &= 0, & x &\in \partial\Omega,
\end{aligned} \]

where \( \Omega \) is a bounded domain in \( \mathbb{R}^N \) with a Lipschitz boundary \( \partial\Omega \), \( a \in L^\infty(\Omega) \) and \( a(x) \geq 0 \) (\( x \in \Omega \)), and \( f : \Omega \times \mathbb{R} \to \mathbb{R} \) satisfies the following standard hypotheses [25]:

(1) \( f(x, \xi) \) is locally Lipschitz on \( \tilde{\Omega} \times \mathbb{R} \).

(2) There is a constant \( c > 0 \) such that \( |f(x, \xi)| \leq c(1 + |\xi|^{s-1}) \) for some \( s \in (2, 2^*) \), where \( 2^* := 2N/(N - 2) \) if \( N \geq 3 \), and \( 2^* := \infty \) if \( N = 1, 2 \).

(3) There are constants \( \mu > 2 \) and \( \delta > 0 \) such that for \( |\xi| \geq \delta \), \( 0 < \mu F(x, \xi) \leq f(x, \xi) \xi \), where \( F(x, \xi) = \int_0^\xi f(x, t)\,dt \).

(4) \( f(x, \xi) = o(\xi) \) as \( \xi \to 0 \).

Define \( H = H_0^1(\Omega) \) with the a-dependent inner product and the norm respectively as

\[ (u, v)_a = \int_\Omega (\nabla u(x) \cdot \nabla v(x) + a(x)u(x)v(x))\,dx, \quad \|u\|_a = \sqrt{(u, u)_a}. \]

Since \( a(x) \) is nonnegative and uniformly bounded, the norm \( \|\cdot\|_a \) is equivalent to the usual norm in \( H_0^1(\Omega) \), i.e., \( \|u\| = \left( \int_\Omega |\nabla u(x)|^2\,dx \right)^{1/2} \). The variational energy functional associated with the BVP (6.1) is given as

\[ E(u) = \int_\Omega \left( \frac{1}{2} |\nabla u(x)|^2 + a(x)|u(x)|^2 - F(x, u(x)) \right)\,dx = \frac{1}{2} \|u\|_a^2 - \int_\Omega F(x, u(x))\,dx. \]

It is well known that under the hypotheses (1)–(4), \( E \in C^1(H, \mathbb{R}) \) and satisfies the (PS) condition. Each critical point of \( E \) is a weak solution and also a classical solution of the BVP (6.1) (see [25]). In addition, \( u \equiv 0 \) is a local minimizer of \( E \). Moreover, in any finite-dimensional subspace of \( H \), \( E(u) \to -\infty \) uniformly as \( \|u\|_a \to \infty \). Hence, for any finite-dimensional closed subspace \( L \), the peak mapping \( P \) of \( E \) with respect to \( L \) is nonempty. According to [17], we assume, in addition to the hypotheses (1)–(4), that

(5) \( f(x, \xi)/|\xi| \) is increasing with respect to \( \xi \) on \( \mathbb{R}\setminus\{0\} \).

For \( L = \{0\} \), as shown in [17], under the hypotheses (1)–(5), \( E \) has only one local minimizer in any direction, i.e., \( E \) has a unique peak selection \( p(v) = t_vv \) (\( \forall v \in S_H \)) with respect to \( L = \{0\} \). Moreover, there exists a \( \delta > 0 \) such that \( t_v = \|p(v)\| \geq \delta \) for any \( v \in S_H \), which is exactly the separation condition in Theorem 4.4 when \( L = \{0\} \). For any finite-dimensional closed subspace \( L \), the uniqueness of the peak selection of \( E \) with respect to \( L \) implies its continuity. As a result, the unique peak selection \( p \) with respect to \( L = \{0\} \) is continuous on \( S_H \). Moreover, if the conditions (1)–(5) hold and

(6) \( f(x, \xi) \in C^1 \) and there exists a constant \( \tilde{c} > 0 \) such that \( |f_t(x, \xi)| \leq \tilde{c}(1 + |\xi|^{s-2}) \) for \( s \) as specified in (2),

then the unique peak selection \( p \) with respect to \( L = \{0\} \) is \( C^1 \) (see [17]).

Due to the limit of the length of the paper, numerical experiments mainly focus on the following three cases of the BVP (6.1) on 2-dimensional domains (square or dumbbell-shaped domains):

**Case 1** (NLSE in the focusing regime [8]). \( f(x, u) = u^3, \ a(x) = \omega|x|^2 \) with \( \omega > 0 \) and \( |x| := \sqrt{x_1^2 + x_2^2} \) for all \( x = (x_1, x_2) \in \Omega = (-1, 1)^2 \subseteq \mathbb{R}^2 \).

**Case 2** (Hénon equation [5, 18]). \( a(x) = 0, \ f(x, u) = |u|^3 \) (\( \ell \geq 0 \)) and \( \Omega = (-1, 1)^2 \).

**Case 3** (Chandrasekhar equation [5]). \( a(x) = 0, \ f(x, u) = (u^2 + 2u)^{3/2}, \ u \geq 0 \) and \( \Omega \) is a 2-dimensional dumbbell-shaped domain as described later.
It is clear that the hypotheses (f1)–(f6) hold for all the functions \( f \) in Cases 1–3. In addition, our approach is efficient for different domains such as an L-shaped domain, a ball or other complex domains in high dimensions.

In our numerical experiments, the initial ascent direction \( v_0 \) is taken as \( v_0 = \tilde{v}_0/\|\tilde{v}_0\|_a \) with \( \tilde{v}_0 \) the solution of the Poisson problem

\[
\begin{cases}
-\Delta \tilde{v}_0(x) = 1_{\Omega_1}(x) - 1_{\Omega_2}(x), & x \in \Omega, \\
\tilde{v}_0(x) = 0, & x \in \partial \Omega,
\end{cases}
\]  

(6.3)

where \( 1(\cdot) \) is the indicator function and \( \Omega_1 \) and \( \Omega_2 \) are two selected disjoint subdomains of \( \Omega \) to control the convexity of \( v_0 \). A numerical solution is reached at \( w_k = p(v_k) \) by the NWP-LMM when

\[
\|g_k\|_a = \|\nabla E(w_k)\|_a \leq 10^{-5}
\]

and

\[
\max_{x \in \Omega} |\Delta w_k(x) - a(x)w_k(x) + f(x, w_k(x))| \leq 5 \times 10^{-5}.
\]

Particularly, it is necessary to explain more about how to numerically compute the gradient of the energy functional and a peak selection in numerical experiments. According to the definition of the gradient \( g_k = \nabla E(w_k) \) of \( E \) at \( w_k \in H \) in (5.1) and the definition of the inner product \( \langle \cdot, \cdot \rangle_a \) in (6.2), the gradient \( g_k \) can be expressed as \( g_k = w_k - \phi_k \) with \( \phi_k \) determined by the linear elliptic BVP

\[
\begin{cases}
-\Delta \phi_k(x) + a(x)\phi_k(x) = f(x, w_k(x)), & x \in \Omega, \\
\phi_k(x) = 0, & x \in \partial \Omega,
\end{cases}
\]  

(6.4)

which can be solved efficiently by a standard numerical method, such as the finite element method (FEM) and the finite difference method. In our numerical code, \texttt{assemde}, an FEM-based subroutine provided by the MATLAB PDE Toolbox, is implemented to accomplish this task. The square domain in Cases 1 and 2 and the dumbbell-shaped domain in Case 3 are, respectively, discretized with 32,768 and 15,552 triangular elements. In addition, the computation of a peak selection \( p(v_k) \) of \( E \) at \( v_k \in S_H \) with respect to a given \((n-1)\)-dimensional closed subspace \( L(n \geq 1) \) is an optimization problem in the \( n \)-dimensional half subspace \([L, v_k] \). To do this, a MATLAB subroutine \texttt{fminunc} with the termination tolerance on the first-order optimality \( 10^{-8} \) is called in our numerical code.

Furthermore, for the convenience of numerical comparisons, we introduce the following notations for three different algorithms:

- **SD-Armijo**: the traditional LMM by utilizing the steepest descent direction \( d_k = -g_k \) and the normalized Armijo-type step-size search rule \([18,32]\). At each iterative step of this algorithm, the step-size \( \alpha_k \) is chosen as

\[
\alpha_k = \max_{m \in \mathbb{N}} \{ \lambda \rho^m : E(p(v_k(\lambda \rho^m))) \leq E(p(v_k)) - \sigma \lambda \rho^m t_k \| g_k \|^2 \},
\]  

(6.5)

where the constants \( \sigma \) and \( \lambda \) are fixed as \( \sigma = 10^{-4} \) and \( \lambda = 0.1 \).

- **SD-StrongWolfe**: the NWP-LMM by utilizing the steepest descent direction \( d_k = -g_k \) and the normalized strong Wolfe-Powell-type step-size search rule \((3.12)\) with constants \( \sigma_1 = 10^{-4} \) and \( \sigma_2 = 0.4 \).

- **CG-StrongWolfe**: the NWP-LMM by utilizing the CG-type direction \((5.5)\) with \( \beta_k = \beta_{FR-like} \) given in \((5.8)\) and the normalized strong Wolfe-Powell-type step-size search rule \((3.12)\) with constants \( \sigma_1 = 10^{-4} \) and \( \sigma_2 = 0.4 \).

It is worthwhile to point out that it is generally not feasible to directly combine the CG-type direction and the Armijo-type step-size search rule in the LMM (i.e., \( d_k \) is not guaranteed to be a descent direction) as observed numerically.
6.1 Numerical results for the nonlinear Schrödinger equation

In this subsection, we report numerical results of Case 1, i.e., the NLSE in the focusing regime on the square domain \( \Omega = (-1,1)^2 \) as

\[
\begin{cases}
  - \Delta u(x) + \omega |x|^2 u(x) = u^3(x), & x \in \Omega, \\
  u(x) = 0, & x \in \partial \Omega.
\end{cases}
\] (6.6)

Taking \( \omega = 8 \), limited by the paper length, we see that only ten different solutions labeled by \( u_1, u_2, \ldots, u_{10} \) are shown in Figure 4 for their profiles and features. The corresponding information on the support space \( L \), subdomains \( \Omega_1 \) and \( \Omega_2 \) used in (6.3), and energy values of these solutions is listed in Table 1. In addition, numerical comparisons of the SD-StrongWolfe, CG-StrongWolfe and SD-Armijo in terms of CPU times for computing these solutions of the NLSE are presented in Figure 5.

From Figures 4 and 5, Table 1 and additional results not shown here, we observe that three LMMs considered are effective for finding multiple solutions of the NLSE in the focusing regime and the CG-StrongWolfe shows the best performance among its LMM companions. In addition, \( u_1 \) (see Figure 4(a)) is the only positive solution with the lowest energy value, i.e., it is the ground state solution.

![Figure 4](image)

**Figure 4** (Color online) Profiles of ten different solutions of the NLSE: (a) the single-peak positive (ground state) solution \( u_1 \) concentrated mainly on the center of the domain; (b)–(j) nine multi-peak sign-changing solutions \( u_2, \ldots, u_{10} \). In each subplot, the horizontal and vertical coordinates measure \( x_1 \) and \( x_2 \), respectively.

| \( u_n \) | \( E(u_n) \) | \( L \) | \( \Omega_1 \) (\( \Omega_2 = \Omega_1 \Omega_1 \)) | Graphics |
|---|---|---|---|---|
| \( u_1 \) | 14.7889 | \{0\} | \( \Omega \) | Figure 4(a) |
| \( u_2 \) | 73.8223 | \{u_1\} | \{x_1 > 0\} | Figure 4(b) |
| \( u_3 \) | 73.8223 | \{u_1\} | \{x_2 > 0\} | Figure 4(c) |
| \( u_4 \) | 70.9151 | \{u_1\} | \{x_1 + x_2 > 0\} | Figure 4(d) |
| \( u_5 \) | 70.9151 | \{u_1\} | \{x_1 - x_2 > 0\} | Figure 4(e) |
| \( u_6 \) | 210.0238 | \{u_1, u_2\} | \{|x_1| > 0.2\} | Figure 4(f) |
| \( u_7 \) | 178.2474 | \{u_1, u_4\} | \{|x_1 + x_2| > 0.3\} | Figure 4(g) |
| \( u_8 \) | 213.6423 | \{u_1, u_2, u_3\} | \{|x_1| > |x_2|\} | Figure 4(h) |
| \( u_9 \) | 243.2646 | \{u_1, u_4, u_5\} | \{|x_1| > |x_2|\} | Figure 4(i) |
| \( u_{10} \) | 306.4755 | \{u_1, u_2, u_3, u_8\} | \{x_1^2 + x_2^2 > 0.25\} | Figure 4(j) |
6.2 Numerical results for the Hénon equation

Now, we report numerical results of Case 2, i.e., the Hénon equation on the square domain \( \Omega = (-1, 1)^2 \) as
\[
\begin{align*}
-\Delta u(x) &= |x|^\ell u(x), & x &\in \Omega, \\
u(x) &= 0, & x &\in \partial\Omega.
\end{align*}
\] (6.7)

Fix \( \ell = 6 \), due to the space limitation, only twelve solutions we obtain, labeled by \( u_1, u_2, \ldots, u_{12} \), are displayed in Figure 6 for their profiles and features. The corresponding information on the support space \( L \), subdomains \( \Omega_1 \) and \( \Omega_2 \) used in (6.3), and energy values of these solutions is listed in Table 2. In addition, numerical comparisons of the \text{SD-StrongWolfe}, \text{CG-StrongWolfe} and \text{SD-Armijo} in terms of CPU times for computing these solutions of the Hénon equation are provided in Figure 7.

From Figures 6 and 7, Table 2 and additional results not shown here, we observe that three LMMs considered can effectively find multiple solutions of the Hénon equation. As expected, the \text{CG-StrongWolfe} also shows the best performance among its LMM companions. In addition, positive solutions are not unique in the case \( \ell = 6 \).

6.3 Numerical results for the Chandrasekhar equation

Here, we report numerical results of Case 3, i.e., the Chandrasekhar equation as
\[
\begin{align*}
-\Delta u(x) &= (u^2(x) + 2u(x))^{3/2}, & x &\in \Omega, \\
u(x) &= 0, & x &\in \partial\Omega.
\end{align*}
\] (6.8)

We consider a dumbbell-shaped domain \( \Omega \) as depicted in Figure 8(a). It contains a smaller disk centered at \((-1, 0)\) with radius 0.5 and a larger disk centered at \((2, 0)\) with radius 1.0. A corridor of width 0.4, symmetric with respect to the \( x_1 \)-axis, is constructed to link the two disks. In this case, we focus on finding multiple positive solutions. Limited by the length of the paper, we only present seven different positive solutions, labeled by \( u_1, u_2, \ldots, u_7 \), in (b)–(h) of Figure 8 for their profiles and features. The corresponding information on the support space \( L \), subdomains \( \Omega_1 \) and \( \Omega_2 \) used in (6.3), and energy values of these solutions is listed in Table 3. In addition, numerical comparisons of the \text{SD-StrongWolfe}, \text{CG-StrongWolfe} and \text{SD-Armijo} in terms of CPU times for computing these solutions of the Chandrasekhar equation are provided in Figure 9. Finally, the comparison of the computational efficiency of the three LMMs by increasing the elements and then freedoms in the FEM for computing the gradient direction \( g_k \) in the iterations for finding \( u_1 \) are listed in Table 4.
Figure 6 (Color online) Profiles of twelve different solutions of the Hénon equation: (a) the single-peak positive (ground state) solution \( u_1 \) concentrated mainly on the corner; (b) a two-peak positive solution \( u_2 \) concentrated mainly on two adjacent corners; (c) a two-peak positive solution \( u_3 \) concentrated mainly on two diagonal corners; (d) a two-peak sign-changing solution \( u_4 \) concentrated mainly on two adjacent corners; (e) a two-peak sign-changing solution \( u_5 \) concentrated mainly on two diagonal corners; (f) a three-peak positive solution \( u_6 \); (g)–(h) two three-peak sign-changing solutions \( u_7 \) and \( u_8 \); (i) a four-peak positive solution \( u_9 \); (j)–(l) three four-peak sign-changing solutions \( u_{10}, u_{11}, u_{12} \). In each subplot, the horizontal and vertical coordinates measure \( x_1 \) and \( x_2 \), respectively.

Table 2 The information on the corresponding solutions of the Hénon equation in Figure 6

| \( u_n \) | \( E(u_n) \) | \( L \) | \( \Omega_1 \) | \( \Omega_2 \) | Graphics |
|---|---|---|---|---|---|
| \( u_1 \) | 61.9634 | \{0\} | \( \{x_1 > 0, x_2 > 0\} \) | \( \emptyset \) | Figure 6(a) |
| \( u_2 \) | 120.7887 | \( [u_1] \) | \( \{x_1 < 0, x_2 > 0\} \) | \( \emptyset \) | Figure 6(b) |
| \( u_3 \) | 122.4078 | \( [u_1] \) | \( \{x_1 < 0, x_2 < 0\} \) | \( \emptyset \) | Figure 6(c) |
| \( u_4 \) | 126.6988 | \( [u_1] \) | \( \{x_2 > 0\} \) | \( \emptyset \) | Figure 6(d) |
| \( u_5 \) | 125.3561 | \( [u_1] \) | \( \{x_1 > 0, x_2 > 0\} \) | \( \{x_1 < 0, x_2 < 0\} \) | Figure 6(e) |
| \( u_6 \) | 177.6068 | \( [u_1, u_2] \) | \( \{x_1 < 0, x_2 < 0\} \) | \( \emptyset \) | Figure 6(f) |
| \( u_7 \) | 187.1379 | \( [u_1, u_3] \) | \( \{x_2 > 0\} \) | \( \emptyset \) | Figure 6(g) |
| \( u_8 \) | 189.9406 | \( [u_1, u_4] \) | \( \{x_1 < 0, x_2 < 0\} \) | \( \emptyset \) | Figure 6(h) |
| \( u_9 \) | 230.0141 | \( [u_1, u_2, u_6] \) | \( \{x_1 > 0, x_2 < 0\} \) | \( \emptyset \) | Figure 6(i) |
| \( u_{10} \) | 247.0220 | \( [u_1, u_2, u_6] \) | \( \{x_2 < 0\} \) | \( \{x_2 > 0\} \) | Figure 6(j) |
| \( u_{11} \) | 250.6746 | \( [u_1, u_2, u_6] \) | \( \{x_1, x_2 > 0\} \) | \( \{x_1 x_2 < 0\} \) | Figure 6(k) |
| \( u_{12} \) | 255.9728 | \( [u_1, u_2, u_6] \) | \( \{x_1 x_2 < 0\} \) | \( \emptyset \) | Figure 6(l) |
Figure 7 (Color online) Comparison of CPU time of LMMs for finding solutions of the Hénon equation in Figure 6

Figure 8 (Color online) (a) A dumbbell-shaped domain. (b)–(h) Profiles of seven different positive solutions of the Chandrasekhar equation: (b) the single-peak positive ground state solution $u_1$ concentrated mainly on the larger disk; (c) a single-peak positive solution $u_2$ concentrated mainly on the smaller disk; (d) a single-peak positive solution $u_3$ concentrated mainly on the corridor; (e)–(g) three two-peak positive solutions $u_4$, $u_5$, $u_6$; (h) a three-peak positive solution $u_7$. In each subplot, the horizontal and vertical coordinates measure $x_1$ and $x_2$, respectively.
Table 3  The information on the corresponding solutions of the Chandrasekhar equation in Figure 8

| $u_n$ | $E(u_n)$ | $L$ | $\Omega_1 \ (\Omega_2 = \emptyset)$ | Graphics |
|-------|----------|-----|------------------------------------|----------|
| $u_1$ | 1.6624   | $\{0\}$ | $(x_1 - 2)^2 + x_2^2 < 1$ | Figure 8(b) |
| $u_2$ | 18.0067  | $\{0\}$ | $(x_1 + 1)^2 + x_2^2 < 0.5$ | Figure 8(c) |
| $u_3$ | 108.0580 | $\{0\}$ | $(x_1 - 0.25)^2 + x_2^2 < 0.1$ | Figure 8(d) |
| $u_4$ | 19.6691  | $[u_1]$ | $(x_1 + 1)^2 + x_2^2 < 0.5$ | Figure 8(e) |
| $u_5$ | 109.6897 | $[u_1]$ | $(x_1 - 0.25)^2 + x_2^2 < 0.1$ | Figure 8(f) |
| $u_6$ | 125.8846 | $[u_2]$ | $(x_1 - 0.25)^2 + x_2^2 < 0.1$ | Figure 8(g) |
| $u_7$ | 127.5247 | $[u_1, u_2]$ | $(x_1 - 0.25)^2 + x_2^2 < 0.1$ | Figure 8(h) |

From Figures 8 and 9, Tables 3 and 4 and additional results not shown here, we observe that three LMMs considered can effectively find multiple positive solutions of the Chandrasekhar equation. It is also observed that the FEM mesh size has little effect on the number of iterations of the three algorithms. Again, the CG-StrongWolfe also shows the best performance in this case.

Above all, numerical experiments in this subsection indicate that the CG-type direction indeed speeds up the LMM greatly.

Table 4  Comparison of the number of iterations (#its) and CPU time (Time) in seconds of LMMs with different numbers of triangular elements ($n_T$) used in the FEM for computing the gradient direction $g_k$ in the iterations for finding the solution $u_1$ of the Chandrasekhar equation in Figure 8

| $n_T$ | SD-Armijo | SD-StrongWolfe | CG-StrongWolfe |
|-------|-----------|----------------|----------------|
|       | #its | Time | #its | Time | #its | Time |
| 3,744 | 51   | 0.4329 | 13   | 0.1270 | 10   | 0.0978 |
| 6,050 | 51   | 0.5782 | 13   | 0.1704 | 9    | 0.1314 |
| 9,732 | 51   | 0.8287 | 13   | 0.2540 | 9    | 0.1757 |
| 15,552 | 51   | 1.2944 | 13   | 0.3921 | 9    | 0.2843 |
| 55,472 | 51   | 4.2249 | 13   | 1.2340 | 9    | 0.9055 |
| 152,696 | 51 | 12.0180 | 13   | 3.3641 | 9    | 2.4026 |
| 635,658 | 51 | 72.2960 | 13   | 19.6600 | 9    | 14.2580 |

Figure 9  (Color online) Comparison of CPU time of LMMs for finding solutions of the Chandrasekhar equation in Figure 8
7 Conclusions

In this paper, we introduced a framework of the normalized Wolfe-Powell-type local minimax method (NWP-LMM) based on general descent directions and the normalized Wolfe-Powell-type and strong Wolfe-Powell-type step-size search rules for finding multiple unstable solutions of semilinear elliptic problems. Under certain conditions on the local peak selection and general descent directions, the feasibility and global convergence of the NWP-LMM were rigorously verified at the functional analysis level. In addition, two feasible types of descent directions, i.e., the preconditioned steepest descent direction and the conjugate-gradient-type direction, were proposed and discussed. The global convergence of the NWP-LMM combined with the preconditioned steepest descent direction was also provided. Extensive numerical results for several semilinear elliptic equations, including the nonlinear Schrödinger equation, the Hénon equation and the Chandrasekhar equation in 2 dimensions, were reported with their multiple solutions displayed to illustrate the effectiveness and robustness of our approach. The superior numerical performance of the NWP-LMM combined with the conjugate-gradient-type direction was observed in extensive numerical experiments, while the rigorous verification for its global convergence is ongoing. Furthermore, designing more efficient preconditioned steepest descent or preconditioned conjugate-gradient-type directions within the framework of the LMM by constructing appropriate preconditioners to further improve the efficiency of computing multiple solutions will be our future work. Finally, it is worthwhile to point out that following the lines of our approach, the steepest descent direction can be replaced by a general descent direction in the device of the traditional normalized Armijo-type and Goldstein-type local minimax algorithms, and both their feasibility and global convergence can be verified.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 12171148 and 11771138) and the Construct Program of the Key Discipline in Hunan Province. Wei Liu was supported by National Natural Science Foundation of China (Grant Nos. 12101252 and 11971007). Wenfan Yi was supported by National Natural Science Foundation of China (Grant No. 11901185), National Key Research and Development Program of China (Grant No. 2021YFA1001300) and the Fundamental Research Funds for the Central Universities (Grant No. 531118010207).

References

1 Al-Baali M. Descent property and global convergence of the Fletcher-Reeves method with inexact line search. IMA J Numer Anal, 1985, 5: 121–124
2 Chang K-C. Infinite Dimensional Morse Theory and Multiple Solution Problems. Boston: Birkhäuser, 1993
3 Chen C M, Xie Z Q. Search extension method for multiple solutions of a nonlinear problem. Comput Math Appl, 2004, 47: 327–343
4 Chen C M, Xie Z Q. Analysis of search-extension method for finding multiple solutions of nonlinear problem. Sci China Ser A, 2008, 51: 42–54
5 Chen G, Zhou J X, Ni W-M. Algorithms and visualization for sign-changing solutions of semilinear elliptic equations. Internat J Bifur Chaos, 2000, 10: 1565–1612
6 Choi Y S, McKenna P J. A mountain pass method for the numerical solution of semilinear elliptic problems. Nonlinear Anal, 1993, 20: 417–437
7 Dai Y H, Yuan Y-X. Nonlinear Conjugate Gradient Methods (in Chinese). Shanghai: Shanghai Scientific & Technical Publishers, 2000
8 Dalfovo F, Giorgini S, Pitaevskii L P, et al. Theory of Bose-Einstein condensation in trapped gases. Rev Modern Phys, 1999, 71: 463–512
9 Ding Z H, Costa D, Chen G. A high-linking algorithm for sign-changing solutions of semilinear elliptic equations. Nonlinear Anal, 1999, 38: 151–172
10 E W N, Ren W Q, Vanden-Eijnden E. String method for the study of rare events. Phys Rev B, 2002, 66: 052301
11 E W N, Zhou X. The gentlest ascent dynamics. Nonlinearity, 2011, 24: 1831–1842
12 Fletcher R. Practical Methods of Optimization. Chichester: John Wiley & Sons, 1987
13 Fletcher R, Reeves C M. Function minimization by conjugate gradients. Comput J, 1964, 7: 149–154
14 Hager W W, Zhang H C. A survey of nonlinear conjugate gradient methods. Pac J Optim, 2006, 2: 35–58
Henkelman G, Jónsson H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. J Chem Phys, 1999, 111: 7010–7022

Le A, Wang Z-Q, Zhou J X. Finding multiple solutions to elliptic PDE with nonlinear boundary conditions. J Sci Comput, 2013, 56: 591–615

Li Y X, Zhou J X. A minimax method for finding multiple critical points and its applications to semilinear PDEs. SIAM J Sci Comput, 2001, 23: 840–865

Li Y X, Zhou J X. Convergence results of a local minimax method for finding multiple critical points. SIAM J Sci Comput, 2002, 24: 865–885

Li Z X, Wang Z-Q, Zhou J X. A new augmented singular transform and its partial Newton-correction method for finding more solutions. J Sci Comput, 2017, 71: 634–650

Liu W, Xie Z Q, Yi W F. Normalized Goldstein-type local minimax method for finding multiple unstable solutions of semilinear elliptic PDEs. Commun Math Sci, 2021, 19: 147–174

Liu W, Xie Z Q, Yuan Y J. Convergence analysis of a spectral-Galerkin-type search extension method for finding multiple solutions to semilinear problems (in Chinese). Sci Sin Math, 2021, 51: 1407–1431

Liu W, Xie Z Q, Yuan Y J. A constrained gentlest ascent dynamics and its applications to finding excited states of Bose-Einstein condensates. J Comput Phys, 2023, 473: 111719

Nocedal J, Wright S J. Numerical Optimization, 2nd ed. New York: Springer, 2006

Powell M J D. Some global convergence properties of a variable metric algorithm for minimization without exact line searches. In: Nonlinear Programming. Proceedings of Symposia in Applied Mathematics, vol. 9. Providence: Amer Math Soc, 1976, 53–72

Ren W Q, Vanden-Eijnden E. A climbing string method for saddle point search. J Chem Phys, 2013, 138: 134105

Sun W Y, Yuan Y-X. Optimization Theory and Methods: Nonlinear Programming. New York: Springer, 2006

Wolfe P. Convergence conditions for ascent methods. SIAM Rev, 1969, 11: 226–235

Wolfe P. Convergence conditions for ascent methods. II: Some corrections. SIAM Rev, 1971, 13: 185–188

Xie Z Q, Chen C M, Xu Y. An improved search-extension method for computing multiple solutions of semilinear PDEs. IMA J Numer Anal, 2005, 25: 549–576

Xie Z Q, Yi W F, Zhou J X. An augmented singular transform and its partial Newton method for finding new solutions. J Comput Appl Math, 2015, 286: 145–157

Xie Z Q, Yuan Y J, Zhou J X. On finding multiple solutions to a singularly perturbed Neumann problem. SIAM J Sci Comput, 2012, 34: A395–A420

Yao X D, Zhou J X. A minimax method for finding multiple critical points in Banach spaces and its application to quasi-linear elliptic PDE. SIAM J Sci Comput, 2005, 26: 1796–1809

Yin J Y, Huang Z, Zhang L. Constrained high-index saddle dynamics for the solution landscape with equality constraints. J Sci Comput, 2022, 91: 62

Yin J Y, Yu B, Zhang L. Searching the solution landscape by generalized high-index saddle dynamics. Sci China Math, 2021, 64: 1801–1816

Yin J Y, Zhang L, Zhang P W. High-index optimization-based shrinking dimer method for finding high-index saddle points. SIAM J Sci Comput, 2019, 41: A3576–A3595

Zhang J Y, Du Q. Shrinking dimer dynamics and its applications to saddle point search. SIAM J Numer Anal, 2012, 50: 1899–1921

Zhang L, Ren W Q, Samanta A, et al. Recent developments in computational modelling of nucleation in phase transformations. NPJ Comput Mater, 2016, 2: 16003

Zhou J X. Solving multiple solution problems: Computational methods and theory revisited. Commun Appl Math Comput, 2017, 31: 1–31