Science & Technology Facilities Council

Technical Report
RAL-TR-2014-006

An efficient dimer method with preconditioning and linesearch

NIM Gould, C Ortner, D Packwood

July 2014

©2014 Science and Technology Facilities Council

This work is licensed under a Creative Commons Attribution 3.0 Unported License.

Enquiries concerning this report should be addressed to:

RAL Library
STFC Rutherford Appleton Laboratory
Harwell Oxford
Didcot
OX11 0QX

Tel: +44(0)1235 445384
Fax: +44(0)1235 446403
email: libraryral@stfc.ac.uk

Science and Technology Facilities Council reports are available online at: http://epubs.stfc.ac.uk

ISSN 1358-6254

Neither the Council nor the Laboratory accept any responsibility for loss or damage arising from the use of information contained in any of their reports or in any communication about their tests or investigations.

A efficient dimer method with preconditioning and linesearch[1]

Nicholas I. M. Gould[2,3], Christoph Ortner[4,5] and David Packwood[4,5]

ABSTRACT

The dimer method is a Hessian-free algorithm for computing saddle points. We augment the method with a linesearch mechanism for automatic step size selection as well as preconditioning capabilities. We prove local linear convergence. A series of numerical tests demonstrate significant performance gains.

---

[1] Issued as arXiv:1407.2817 on July 10th, 2014.

[2] Computational Science and Engineering Department, Rutherford Appleton Laboratory, Chilton, Oxfordshire, OX11 0QX, England, EU. Email: nick.gould@stfc.ac.uk . Current reports available from "http://www.numerical.rl.ac.uk/reports/reports.shtml".

[3] The work of this author was supported by the EPSRC grant EP/J021377/1.

[4] Mathematics Institute, Zeeman Building, University of Warwick, Coventry CV4 7AL, England, EU. Email: christoph.ortner@warwick.ac.uk, d.packwood@warwick.ac.uk . Current reports available from "http://warwick.ac.uk/staff/C.Ortner".

[5] This work of these authors was supported by the EPSRC grant EP/J022055/1..

Scientific Computing Department
Rutherford Appleton Laboratory
Oxfordshire OX11 0QX

10 July 2014

1 Introduction

The problem of determining saddle points on high dimensional surfaces has received a great deal of attention from the chemical physics community over the past few decades. These surfaces arise, in particular, as potential energies of molecules or materials. The local minima of such functions describe stable atomistic configurations, while saddle points provide information about the transition rates between minima in the harmonic approximation of transition state theory. Independently, they are useful for mapping the energy landscape and are used to inform accelerated MD type schemes such as hyperdynamics [24, 22] or kinetic Monte Carlo (KMC) [25].

While the problem of determining the minima of such an energy function is well known in the numerical analysis community, the problem of locating saddles point has received little attention. Saddle search algorithms can be broadly categorised into two groups.

The first group has been called 'chain of states' methods. A chain of 'images' are placed on the energy surface, often the two end points of the chain are placed at two different local minima, for which the connecting saddle is being sought. The chain is then 'relaxed' by some dynamics for which the mininum energy path (MEP) is (thought to be) an attractor. Two archetypical methods of this class are the nudged elastic band (NEB) method [11] and the string method [26, 27].

The second group of methods for finding the saddle have been called 'walker' methods. Here a single 'image' moves from its initial point (sometimes, but not obligatorily, a local minimum) until it becomes sufficiently close to a saddle point. The first method to work in this framework was Rational Function Optimization (RFO) and later its derivative, the Partitioned RFO (PRFO) [7, 21, 3]. Here, the full eigenstructure of the Hessian is explicitly calculated and then one or more eigenvalues are manually shifted. In particular, if the minimum eigenvalue is shifted in the correct manner, and a Newton step is applied using the resultant modified Hessian, then the walker moves uphill in the direction corresponding to the lowest eigenvector and downhill in all other directions. If the Hessian is expensive to calculate, or even unavailable, it can be approximated as the computation proceeds by any variety of techniques, for example the symmetric rank-one approximation [18]. Of course any useful Hessian approximation should necessarily have the flexibility to be indefinite. Other walker type techniques are satisfied with computing the lowest eigenpair only. One such technique is the Activation Relaxation Technique (ART) nouveau [16, 15, 17, 6]. The original ART method used an ascent step not along the minimum eigenvector, but along a line drawn between the image and a known local minimum [4, 5]. In ART nouveau this is replaced by the minimum eigenpair which is calculated by means of the Lanczos [13] method.

The technique which forms the basis of the present paper, is the *dimer method* [9, 10]. In this method a pair of 'walkers' is placed on the energy surface and aligned with the minimum eigenvector (irrespective of the sign of the corresponding eigenvalue) by minimizing the sum of the energies at the two end points. This can be thought of as the computation of the minimal eigenvalue using a finite difference approximation to the

Hessian matrix. In practice this 'rotation step' is not converged to great precision. More advanced modifications can be used to improve walker search directions, e.g., an L-BFGS [14] scaling, rather than just using a default steepest descent type scheme [12].

In the only rigorous analysis of the dimer method that we are aware of Zhang and Du [28] prove local convergence of a variation where the 'dimer length' (the separation distance between the two walkers) shrinks to zero. In that work the dimer evolution is treated as a dynamical system, and the stability of different types of equilibria is investigated.

In the present paper we present three new results:

1. We augment the dimer method with preconditioning capabilities to improve its efficiency for ill-conditioned problems, in particular with an eye to high-dimensional molecular energy landscapes. This modification is based on the elementary observation that the dimer method can be formulated with respect to an arbitrary inner product. (The ℓ^2-inner product was previously used exclusively.)

2. We introduce a linesearch procedure. To that end, the main difficulty is the absence of a merit function for saddles. Instead, we proposed a *local merit function*, which we minimise at each dimer iteration using traditional linesearch strategies from optimisation, and which is updated between steps.

3. We present a variation of the analysis of Zhang and Du [28] that demonstrates that it is unnecessary to shrink the dimer length, h, to zero. Indeed, shrinking h can cause severe numerical difficulties due to round-off. We prove that, if it is kept fixed, then the dimer walkers converge to a point that lies within $O(h^2)$ of a saddle. We also extend this analysis to incorporate preconditioning and linesearch.

Concerning (2), it would of course be preferable to construct a global merit function as this would provide a path towards constructing a globally convergent scheme. Indeed, our (non-trivial) generalisation of the convergence analysis to the linesearch variant of the dimer method only yields local results, and we even present counterexamples to global convergence.

The paper is organised as follows: having established preliminary concepts, we describe two variants of the basic dimer method, and establish their local convergence, in §2. A linesearch enhancement is proposed, and its local convergence behaviour is analysed, in §3. Numerical experiments illustrating the advantages of the linesearch are given in §4. We conclude in §5. Full details of our analysis are given in Appendix A.

2 Local Convergence of the Dimer Method

2.1 Preliminaries

Let X be a Hilbert space with norm $\|x\|$ and inner product $x \cdot y$. We write $x \perp y$ if $x \cdot y = 0$. $I : X \to X$ denotes the identity. For $x, y \in X$, $x \otimes y : X \to X$ denotes the operator defined by $(x \otimes y)z = (y \cdot z)x$.

Given two real functions f and g defined in some neighbourhood $\mathcal{N}$ of the origin, we say that $f(x) = O(g(x))$ as $x \to 0$ if $|f(x)| \leq C|g(x)|$ for some constant $C > 0$ and all $x \in \mathcal{N}$.

For a bounded linear operator $A \in L(X)$ we denote its spectrum by $\sigma(A)$. We say that $(\lambda, v) \in \mathbb{R} \times X$ is an eigenpair if $Av = \lambda v$. If (λ, v) is an eigenpair and $\lambda = \inf \sigma(A)$, then we call it a minimal eigenpair. We say that A has *index-1 saddle structure* if there exists a unique minimal eigenpair (λ, v) with $\lambda < 0$ and A is positive definite in $\{v\}^\perp$.

If $F : X \to \mathbb{R}$ is Fréchet differentiable at a point x then we denote its *gradient* by $\nabla F(x)$, i.e.,

$$\nabla F(x) \cdot y = \lim_{t \to 0} t^{-1}(F(x + ty) - F(x)).$$

(Note that $\nabla F(x)$ is the Riesz representation of the first variation $\delta F(x) \in X^*$.) Similarly, if $F : X \to X$ is Fréchet differentiable at x, then $\nabla F(x) \in L(X)$ is a bounded linear operator satisfying $\nabla F(x)u = \lim_{t \to 0} t^{-1}(F(x + tu) - F(x))$. In particular, if $F : X \to \mathbb{R}$, then the Hessian $\nabla^2 F(x) \in L(X)$ (rather than $\nabla^2 F(x) : X \to X^*$). Higher derivatives are defined analogously, but we shall avoid their explicit use as much as possible.

We say that x_* is an *index-1 saddle* of E if

$$\nabla E(x_*) = 0 \quad \text{and} \quad \nabla^2 E(x_*) \text{ has index-1 saddle structure.} \tag{1}$$

With slight abuse of notation, we shall also call (x_*, v_*, λ_*) an index-1 saddle if x_* is an index-1 saddle and (v_*, λ_*) the associated minimal eigenpair.

Given a dimer length h and a vector $v \in S_1 := \{u \in X \mid \|u\| = 1\}$, we define

$$\begin{aligned} \mathcal{E}_h(x, v) &:= \tfrac{1}{2}\big(E(x + hv) + E(x - hv)\big) \quad \text{and} \\ E_h(x) &:= \inf_{v \in S_1} \mathcal{E}_h(x, v). \end{aligned}$$

If $\# \arg\min_{v \in S_1} \mathcal{E}_h(x, v) = 1$, then we also define

$$V(x) := \arg\min_{v \in S_1} \mathcal{E}_h(x, v)$$

and we can then write $E_h(x) = \mathcal{E}_h(x, V(x))$.

Finally, we observe that

$$\nabla_v \mathcal{E}_h(x, v) = \tfrac{h}{2}\big(\nabla E(x + hv) - \nabla E(x - hv)\big) = h^2 \nabla^2 E(x)v + O(h^4) \quad \text{and} \tag{2}$$

$$\nabla_v^2 \mathcal{E}_h(x, v) = \tfrac{h^2}{2}\big(\nabla^2 E(x + hv) + \nabla^2 E(x - hv)\big) = h^2 \nabla^2 E(x) + O(h^4), \tag{3}$$

and we therefore define the discrete Hessian operator

$$H_h(x; v) := h^{-2} \nabla_v \mathcal{E}_h(x, v). \tag{4}$$

2.2 Two basic dimer variants

We now make precise two basic variants of the dimer method. The first algorithm is a variation of the original dimer method [9, 20], alternating steps in the position (x_k) and direction (s_k) variables, but employs a modification proposed by [28]. Indeed, the following algorithm can be thought of as [28] with λ (h in our case) taken to be constant instead of $h \to 0$ as $k \to \infty$.

Algorithm 1

(0) Choose $x_0, v_0 \in X$ with $\|v_0\| = 1$, $h > 0$ and step lengths $(\alpha_k)_{k\in\mathbb{N}}, (\beta_k)_{k\in\mathbb{N}}$.
(1) For $n = 0, 1, 2, \ldots$ do
(2) $\quad s_k := -(I - v_k \otimes v_k) h^{-2} \nabla_v \mathcal{E}_h(x_k, v_k)$
(3) $\quad v_{k+1} := \cos(\|s_k\|\beta_k) v_k + \sin(\|s_k\|\beta_k) \frac{s_k}{\|s_k\|}$
(4) $\quad x_{k+1} := x_k - \alpha_k (I - 2v_k \otimes v_k) \nabla_x \mathcal{E}_h(x_k, v_k)$.

Our second variant of the dimer method that we consider is closer in spirit to the class of walking methods which employ the minimal eigenpair. These include Rational Function Optimization (RFO) [7, 21, 3], which uses either an exact or approximate Hessian directly, or the Activation Relaxation Technique nouveau (ART Nouveau)[16, 15, 17], which uses the Lanczos method to find the minimal eigenvector. This modification of the dimer method can also be motivated by observations in [20] that undertaking more accurate rotation steps may lead to fewer iterations. As an idealised variant of this idea we consider a dimer algorithm where, at each iteration, an exact rotation v is computed.

Algorithm 2

(0) Choose $x_0 \in X, h > 0, (\alpha_k)_{k\in\mathbb{N}}$.
(1) For $k = 0, 1, 2, \ldots$ do
(2) $\quad v_k \in \arg\min_{\|v\|=1} \mathcal{E}_h(x_k, v)$
(3) $\quad x_{k+1} = x_k - \alpha_k (I - 2v_k \otimes v_k) \nabla_x \mathcal{E}_h(x_k, v_k)$

Remark 1. 1. Algorithm 1 is clearly well-defined. Algorithm 2 is well-defined if $\dim(X) < \infty$, however, step (2) in Algorithm 2 is not necessarily well-defined in Hilbert space. We shall show in Theorem 3(b) that this step is well-defined if the starting guess is close to a saddle point. In practice, the minimisation with respect to v may only be performed to within a specified tolerance (see §3.2).

2. Both Algorithm 1 and Algorithm 2 may be rewritten such that a step in the position variable x is performed by employing the gradient $\nabla E(x)$ instead of the averaged gradient $\nabla_x \mathcal{E}_h(x_k, v_k)$. For the sake of uniformity and simplicity of presentation we do not explicitly consider these as well.

However, we note that (1) all our results can be extended to these variants, and (2) it seems to us that this has minor effects on the accuracy and efficiency of the algorithms, with the exception that it requires additional gradient evaluations.

Instead, it might be advisable to "post-process" the dimer Algorithms 1 and 2 using such a modified scheme. Namely, we shall prove that Algorithms 1 and 2 converge to a point (x_h, v_h) that is $O(h^2)$ close to an index-1 saddle. Post-processing would then yield the exact saddle point.

3. A natural variant of step (4) of Algorithm 1 is to replace it with

$$x_{k+1} := x_k - \alpha_k(I - 2v_{k+1} \otimes v_{k+1})\nabla_x \mathcal{E}_h(x_k, v_{k+1}).$$

We have observed that, in practise, this does not change the number of iterations required to reach a specified residual, but that it doubles the number of force (gradient) evaluations. Note that with the formulation we use, $\nabla_v \mathcal{E}_h(x_k, v_k) = \frac{1}{2}(\nabla E(x_k + hv_k) - \nabla E(x_k - hv_k))$, and $\nabla_x \mathcal{E}_h(x_k, v_k) = \frac{1}{2}(\nabla E(x_k + hv_k) + \nabla E(x_k + hv_k))$ and therefore only two force evaluations $\nabla E(x_k \pm hv_k)$ are required. The variant proposed in item 2. of the present remark would require three force evaluations in each step. □

2.3 The dimer saddle

Our first observation is that the dimer method (in both variants we consider) approximates the Hessian by a finite difference and the gradient by an average. Therefore, the dimer iterates (x_k, v_k) with fixed dimer length h *cannot* in general converge to a saddle but only to a critical point (x_h, v_h) near a saddle, satisfying

$$\nabla_x \mathcal{E}_h(x_h, v_h) = 0 \;\text{ and }\; (I - v_h \otimes v_h)\nabla_v \mathcal{E}_h(x_h, v_h) = 0. \tag{5}$$

The existence (and local uniqueness) of such critical points is established in the following result.

Proposition 2. *Let (x_*, v_*, λ_*) be an index-1 saddle, then there exists $h_0 > 0$ such that, for all $h \le h_0$, there exist $x_h, v_h \in X$, $\lambda_h \in \mathbb{R}$ and a constant C, such that*

$$\begin{aligned} \nabla_x \mathcal{E}_h(x_h, v_h) &\equiv \tfrac{1}{2}\big(\nabla E(x_h + hv_h) + \nabla E(x_h - hv_h)\big) = 0, \\ \tfrac{1}{h^2}\nabla_v \mathcal{E}_h(x_h, v_h) &\equiv \tfrac{1}{2h}\big(\nabla E(x_h + hv_h) - \nabla E(x_h - hv_h)\big) = \lambda_h v_h, \\ &\qquad\qquad \tfrac{1}{2}\|v_h\|^2 = \tfrac{1}{2}. \end{aligned} \tag{6}$$

and moreover

$$\|x_h - x_*\| + \|v_h - v_*\| + |\lambda_h - \lambda_*| \le Ch^2. \tag{7}$$

Idea of proof. The result is a consequence of the inverse function theorem. Comparing (6) with the exact saddle (x_*, v_*, λ_*) a Taylor expansion shows that the residual is of order $O(h^2)$. Similarly, the linearisation can be shown to be $O(h^2)$ close (in operator norm) to the linearisation of the exact saddle system $\nabla E(x_*) = 0, \nabla^2 E(x_*)v_* = \lambda_* v_*, \|v_*\| = 1$. The linearisation of the latter is an isomorphism by the assumption that x_* is an index-1 saddle. The complete proof is given in A.1. □

We shall refer to a triple $(x_h, v_h, \lambda_h) \in X \times X \times \mathbb{R}$ that satisfies (6) as a *dimer saddle*.

2.4 Local convergence

We now state local convergence results for the two dimer variants formulated in Algorithm 1 and Algorithm 2. The main observation is that Algorithm 1 need not converge monotonically, but that Algorithm 2 is in fact contractive.

Theorem 3. *Let x_* be an index-1 saddle with minimal eigenpair (λ_*, v_*). Then there exists a radius r, a maximal dimer length h_0 and maximal step sizes $\bar\alpha$ and $\bar\beta$ (independent of one another) as well as a dimer saddle (x_h, v_h, λ_h) satisfying (6) such that the following hold for all $h \le h_0$:*

(a) Let $x_0 \in B_r(x_*), v_0 \in B_r(v_*), \sup_k \alpha_k \le \bar\alpha, \sup \beta_k \le \bar\alpha, \inf_k \alpha_k > 0, \inf \beta_k > 0$, and let (x_k, v_k) be the iterates generated by Algorithm 1, then there exist $C > 0, \eta \in (0,1)$ such that

$$\|x_k - x_h\| + \|v_k - v_h\| \le C\eta^k\big(\|x_0 - x_h\| + \|v_0 - v_h\|\big). \tag{8}$$

(b) Let $x_0 \in B_r(x_*), h \le h_0, \sup_k \alpha_k \le \bar\alpha, \inf_k \alpha_k > 0$, then Algorithm 2 is well-defined (i.e., step (2) has a unique solution) and there exists $\eta \in (0,1)$ such that

$$\|x_{k+1} - x_h\| \le \eta \|x_k - x_h\| \qquad \text{for all } k \ge 0. \tag{9}$$

Moreover, there exists a constant C such that $\|v_k - v_h\| \le C\|x_k - x_h\|$.

Idea of proof. (a) The proof of case (a) is a modification of the proofs of [28, Thm. 2.1 and Thm. 3.1]. Upon linearisation of the updates about the exact saddle (x_*, v_*), the updates can be re-written as

$$\begin{pmatrix} x_{k+1} - x_h \\ v_{k+1} - v_h \end{pmatrix} = \left[I - \begin{pmatrix} \alpha_k A & 0 \\ \beta_k B & \beta_k C \end{pmatrix} \right] \begin{pmatrix} x_k - x_h \\ v_k - v_h \end{pmatrix} + O\big((\alpha_k + \beta_k)(h^2 + r_k) r_k\big), \tag{10}$$

where $r_k^2 = \|x_k - x_h\|^2 + \|v_k - v_h\|^2$,

$$A = (I - 2v_* \otimes v_*)\nabla^2 E(x_*), \qquad C = (I - v_* \otimes v_*)\nabla^2 E(x_*) - \lambda_* I, \tag{11}$$

and B is a bounded linear operator (the precise form is not important).

Clearly, A, C are both symmetric and positive definite, hence the spectrum of $\mathbf{A} = (\alpha A, 0; \beta B, \beta C)$ is strictly positive. If we chose $\alpha_k \equiv \alpha, \beta_k \equiv \beta$ constant, then (8) follows from standard stability results for dynamical systems. The (straightforward) generalisation, together with complete proof of (10) are given in §A.2

(b) We first note that step (2) of Algorithm 2 is well-defined due to the fact that $\nabla^2 E(x)$ has index-1 structure for all $x \in B_r(x_*)$, if r is chosen sufficiently small. In this case an implicit function argument guarantees the existence of a unique solution $v_k = V(x_k)$. This is made precise in Lemma 13.

In the same lemma we also show that $\nabla^2 E_h(x) = \nabla^2 E(x) + O(h^2)$ for all $x \in B_r(x_*)$. This allows us to linearize step (3) in Algorithm 2 to obtain

$$x_{k+1} - x_h = \big(I - \alpha_k A\big)(x_k - x_h) + O(r_k^2 + h^2 + \alpha_k^2) r_k,$$

where, again, $A = (I - 2v_* \otimes v_*)\nabla^2 E(x_*)$ and $r_k = \|x_k - x_h\|$. Since A is positive definite the result follows easily. The complete proof is given in §A.3. □

3 A Dimer Algorithm with Linesearch

3.1 Motivation: a local merit function

Let $x_* \in X$ be an index-1 saddle with minimal eigenpair (v_*, λ_*), and consider the modified energy functional

$$F(x) := E(x) + \frac{\kappa}{2}\big(v_* \cdot (x - x_*)\big)^2.$$

Then, $\nabla F(x_*) = 0$ and $\nabla^2 F(x_*) = (I + \kappa v_* \otimes v_*)\nabla^2 E(x_*)$, which is positive definite if and only if $\kappa > -\lambda_*$. For this choice, it follows that x_* is a strict local minimizer of F.

The dimer variant of this observation is that, if (x_h, v_h, λ_h) is a dimer saddle point (cf. Theorem 2) and we define a modified energy functional

$$F_h(x) := \mathcal{E}_h(x, v_h) + \frac{\kappa}{2}\big(v_h \cdot (x - x_h)\big)^2,$$

then choosing $\kappa > -\lambda_*$ and h sufficiently small again guarantees that x_h becomes a local minimizer of F_h. We can make this precise (and generalise) as follows.

Lemma 4. *Let $x_0 \in X$ such that $\nabla^2 E(x_0)$ has index-1 saddle structure with minimal eigenpair $(V(x_0), \lambda)$ and $\mu > 0$ such that $y \cdot (\nabla^2 E(x_0) y) \geq \mu \|y\|^2$ for $y \in \{V(x_0)\}^\perp$. Fix $r, h_0 > 0$.*

Let $0 < h \leq h_0, v_0 \in X, \|v_0\| = 1, g_0 \in X$ and

$$F_0(x) := \mathcal{E}_h(x, v_0) + g_0 \cdot (x - x_0) + \tfrac{\kappa}{2}\big(v_0 \cdot (x - x_0)\big)^2,$$

then there exists $C = C(x_0, r, h_0)$ such that, for all $x \in B_r(x_0), h < h_0, y \in X$,

$$y \cdot (\nabla^2 F_0(x) y) \geq \Big(\min\big(\mu, \kappa + \lambda\big) - C\big(h^2 + \|v_0 - V(x_0)\| + \|x - x_0\|\big)\Big)\|y\|^2.$$

Proof. For $x = x_0$, we compute $\nabla^2 F_0(x_0) = \nabla^2 E(x_0) + O(h^2) + \kappa v_0 \otimes v_0$. Then, the result follows readily from the observation that

$$\begin{aligned}(v_0 \cdot y)^2 &= (v \cdot y)^2 + ((v_0 - v) \cdot y)((v_0 + v) \cdot y) \\ &\geq (v \cdot y)^2 - 2\|v_0 - v\|\|y\|^2.\end{aligned}$$

For general x, the result follows from local Lipschitz continuity of $\nabla^2 E$. □

To complete the definition of F_0 we must specify g_0, κ. The strategy is to choose it in such a way that minimising F_0 will lead to an improved approximation for x.

From the inverse function theorem it follows that there exists $\tilde{x} = x_* + O(h^2) = x_h + O(h^2)$ such that $\nabla_x \mathcal{E}_h(\tilde{x}, v_0) = 0$ (we will make this precise below), and Lemma 4 allows us to assume that it is in fact a local minimiser of F_0. When minimising F_0, we therefore hope to obtain a point "close to" $\tilde{x}$. To test this, we evaluate the residual at $x = \tilde{x}$,

$$\begin{aligned}\nabla F_0(\tilde{x}) &= \nabla_x \mathcal{E}_h(\tilde{x}, v_0) + g_0 + \kappa(v_0 \otimes v_0)(\tilde{x} - x_0) \\ &\approx g_0 + \kappa(v_0 \otimes v_0)\nabla_x^2 \mathcal{E}_h(x_0, v_0)^{-1}\big(\nabla_x \mathcal{E}_h(\tilde{x}, v_0) - \nabla_x \mathcal{E}_h(x_0, v_0)\big) \\ &\approx g_0 - \tfrac{\kappa}{\lambda_0}(v_0 \otimes v_0)\nabla_x \mathcal{E}_h(x_0, v_0),\end{aligned}$$

where $\lambda_0 = H_h(x_0; v_0) \cdot v_0$. This leads to the choice

$$g_0 := \frac{\kappa}{\lambda_0}(v_0 \otimes v_0)\nabla_x \mathcal{E}_h(x_0, v_0).$$

Note in particular, that the steepest descent direction for F_0 at x_0 is

$$-\nabla F_0(x_0) = \big(I + \tfrac{\kappa}{\lambda_0}(v_0 \otimes v_0)\big)\nabla_x \mathcal{E}_h(x_0, v_0).$$

For the special choice $\kappa = -2\lambda_0$, this yields the standard dimer search direction.

3.2 Dimer algorithm with linesearch

Given an iterate x_k, v_k and $\lambda_k := v_k \cdot H_h(x_k; v_k)$, we define the auxiliary functional $F_k \in C^4(X)$,

$$\begin{aligned}F_k(x) &:= \mathcal{E}_h(x, v_k) - 2\big[(v_k \otimes v_k)\nabla_x \mathcal{E}_h(x_k, v_k)\big] \cdot (x - x_k) - \lambda_k \big\|(v_k \otimes v_k)(x - x_k)\big\|^2 \quad (12) \\ &= \mathcal{E}_h(x, v_k) - 2\big((v_k \cdot \nabla_x \mathcal{E}_h(x_k, v_k)\big) \cdot \big(v_k \cdot (x - x_k)\big) - \lambda_k \big(v_k \cdot (x - x_k)\big)^2,\end{aligned}$$

motivated by the discussion in §3.1. Instead of locally minimising F_k we only perform a minimisation step in the steepest descent direction, using a standard linesearch procedure augmented with the following sanity check: For a trial $x^{\mathrm{t}} = x_k - \alpha \nabla F_k(x_k)$ we require that v_k is still a reasonable dimer orientation for x^{t} by checking the residual $\|(I - v_k \otimes v_k)H_h(x^{\mathrm{t}}; v_k)\|$. If this residual falls above a certain tolerance then we reject the step and reduce the step size.

Algorithm 3:

1. **Input:** x_0, v_{-1}, h
   Parameters: $\beta_{-1}, \alpha_0, \alpha_{\max} > 0, \Theta \in (0, 1), \Psi > 1$
2. **For** $k = 0, 1, 2, \ldots$ **do**
   %% Rotation %%
3. $[v_k, \beta_k] := \mathbf{Rotation}[x_k, v_{k-1}, \beta_{k-1}]$
   %% Translation %%
4. $p := -\nabla F_k(x_k)$

5. $\alpha := \min(\alpha_{\max}, 2\alpha_{k-1})$
6. **While** $(F_k(x_k + \alpha p) > F_k(x_k) - \Theta\alpha\|p\|^2)$
   or $(\|(I - v_k \otimes v_k)H_h(x_k + \alpha p; v_k)\| > \Psi\|\nabla_x \mathcal{E}_h(x_k, v_{k-1})\|)$ **do**
7. $\alpha := \alpha/2$
8. $x_{k+1} := x_k + \alpha p$; $\alpha_k := \alpha$

It remains to specify step (3) of Algorithm 3. Any method computing an update v_k satisfying $\|(I - v_k \otimes v_k)H_h(x_k; v_k)\| \leq \mathrm{TOL}$, for given TOL, is suitable; we prescribe the tolerance $\mathrm{TOL} = \|\nabla_x \mathcal{E}_h(x_k, v_{k-1})\|$ so long as this isn't too large. A basic choice of method is the following projected steepest descent algorithm.

Rotation:

1. **Input:** x, v, β
   Parameters: $\mathrm{TOL} = \min(\|\nabla_x \mathcal{E}_h(x, v)\|, \mathrm{TOL}_v^{\mathrm{hi}})$, $\beta_{\max}$, Θ
2. **While** $\|(I - v \otimes v)H_h(x; v)\| > \mathrm{TOL}$ **do**
3. $s := -(I - v \otimes v)H_h(x; v)$
4. $r := \|s\|$; $\beta := \min(\beta_{\max}, 2\beta)$
5. $v_\beta := \cos(\beta r)v + \sin(\beta r)\frac{s}{r}$
6. **While** $\mathcal{E}_h(x, v_\beta) > \mathcal{E}_h(x, v) - \Theta\beta\|s\|^2$ **do**
7. $\beta := \beta/2$
8. $v := v_\beta$
9. **Output:** v, β

Proposition 5. *Algorithm 3 is well-defined in that the rotation step (3) as well as the linesearch loop (6, 7) both terminate after a finite number of iterations, the latter provided that* $\nabla_x \mathcal{E}_h(x_k, v_k) \neq 0$.

Proof. The Rotation Algorithm employed in step (3) of Algorithm 3 terminates for any starting guess due to the fact that it is a steepest descent algorithm on a Stiefel manifold (the unit sphere) with a backtracking linesearch employing the Armijo condition [23]. Convergence of this iteration to a critical point is well known [1, Chap.4]. The loop (6,7) terminates after a finite number of iterations [19] since p is a descent direction for $F_k \in C^4$, that is, $F_k(x_k + \alpha p) = F_k(x_k) - \alpha\|p\|^2 + O(\alpha^2)$. □

Remark 6. 1. In practise, the algorithm terminates, once the entire dimer saddle residual reaches a prescribed tolerance, i.e., $\|\nabla_x \mathcal{E}_h(x_k, v_k)\| \leq \mathrm{TOL}^x$ in addition to $\|(I - v_k \otimes v_k)H_h(x_k; v_k)\| \leq \|\nabla_x \mathcal{E}_h(x_k, v_k)\|$.

2. The two basic backtracking linesearch loops (5)–(8) and (11)–(12) can (and should) be replaced with more effective linesearch routines in practise, in particular choosing more effective starting guesses and using polynomial interpolation to compute linesearch steps. However, the discussion in §3.3 indicates that a Wolfe-type termination criterion might be inappropriate. □

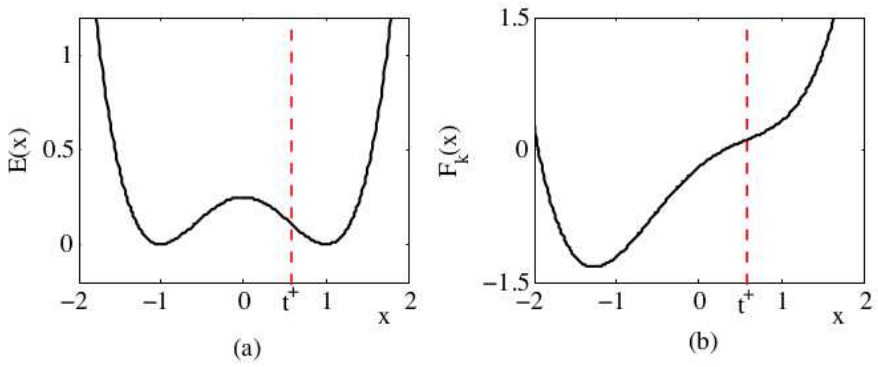

Figure 1: (a) Double-well energy defined in (13). (b) The auxiliary functional $F_k(x)$ with $x_k = t_h^+$; cf. §3.3. The second turning point $t_h^- = -t_h^+$ is an admissible descent step for F_k, hence the dimer method can potentially cycle.

3.3 Failure of global convergence

The modifications of the original dimer algorithms that we have in Algorithm 3 would, in the case of optimisation, yield a globally convergent scheme. Unfortunately, this is not the case in the saddle search case. To see this, consider a one-dimensional double-well example,

$$E(x) = \tfrac{1}{4}(1-x^2)^2 = \tfrac{1}{4}x^4 - \tfrac{1}{2}x^2 + \tfrac{1}{4}; \tag{13}$$

cf. Figure 1(a). There are only two possible (equivalent) dimer orientation $v = \pm 1$, and therefore the rotation steps in Algorithm 3 are ignored. We always take $v = 1$ without loss of generality. The translation search direction at step k is always given by $p = -(1-2)\nabla_x \mathcal{E}_h(x_k, 1) = \nabla_x \mathcal{E}_h(x_k, 1)$, i.e., an ascent direction.

It is easy to see that $x_* = 0$ is an index-1 saddle (i.e., a maximum), and that there are two turning points $t^\pm = \pm 3^{-1/2}$. Thus, there exist "discrete turning points" $t_h^\pm = \pm 3^{-1/2} + O(h^2)$ such that $\lambda(t_h^\pm) = 0$, where $\lambda(x) = H_h(x;1)\cdot 1 = \frac{1}{2h^2}(E'(x+h) - E'(x-h))$.

Suppose that we have an iterate $x_k = t_h^+$, then the translation search direction is $p^+ = \nabla_x \mathcal{E}_h(t_h^+, 1) < 0$. Since $\mathcal{E}_h(t_h^-) = \mathcal{E}_h(t_h^+)$ it follows that

$$F_k(t_h^-) = \mathcal{E}_h(t_h^-) - 2p^+(t_h^- - t_h^+) < \mathcal{E}_h(t_h^-) = F_k(t_h^+).$$

Thus, for Θ sufficiently small, the update $x_{k+1} = t_h^-$ satisfies all the conditions for termination of the loop (11)–(12) in Algorithm 3. See also Figure 1 (b), where F_k is visualised.

We therefore conclude that our newly proposed variant of the dimer algorithm does not excluded cycling behaviour. We also remark that the example is not exclusively one-dimensional, but that analogous constructions can be readily made in any dimension.

3.4 Local convergence

We now establish a local convergence rate.

Theorem 7. *Let (x_*, v_*, λ_*) be an index-1 saddle then there exist $r, h_0, C > 0$ and $\gamma \in (0,1)$ such that, for $x_0 \in B_r(x_*)$, $v_{-1} \in B_r(v_*)$ with $\|v_{-1}\| = 1$ and $h \le h_0$, the iterates x_k, v_k generated by the algorithm satisfy*

$$\|x_k - x_h\| + \|v_k - v_h\| \le C\gamma^k\big(\|x_0 - x_h\| + h^2\|v_{-1} - v_h\|\big), \tag{14}$$

where (x_h, v_h) is the dimer saddle associated with (x_*, v_*, λ_*); cf. Theorem 2.

Sketch of proof. Let $r_k = \|x_k - x_h\|$ and $s_k := \|v_k - v_h\|$.

0. We recall basic contraction results for Armijo-based linesearch methods both in a general Hilbert space and for iterates constrained to lie on the unit sphere in §A.4.

1. As a first proper step we establish that, under the termination criterion $\|(1 - v_k \otimes v_k)H_h(x_k; v_k)\| \le \|\nabla_x \mathcal{E}_h(x_k, v_{k-1})\|$ for the rotation step, it follows that $\|v_k - v_h\| \lesssim r_k + h^2 s_{k-1}$. This is proven in Lemma 16 and Lemma 17.

2. Next, we use this result to establish that there exists a local minimizer y_k of F_k satisfying $\|y_k - x_h\| \lesssim r_k^2 + h^2 r_k + h^4 s_{k-1}$. This is established in Lemma 18.

3. The linesearch procedure and the upper bound on the step length ensure that the step of x_k to x_{k+1} contracts towards y_k, that is, $\|x_{k+1} - y_k\|_* \le \gamma_* \|x_k - y_k\|_*$ for some $\gamma_* \in (0,1)$ and $\|\cdot\|_*$ the *energy* norm induced by $(I - 2v_* \otimes v_*)\nabla^2 E(x_*) \approx \nabla^2 F_k(y_k)$. This is obtained in Lemma 19.

4. The three preceding steps can then be combined to establish that, for r_0, s_{-1}, h sufficiently small, there exists a constant $\gamma_3 \in (\gamma_*, 1)$ such that

$$r_{k+1}^* + h^2 s_k \le \gamma_3 (r_k^* + h^2 s_{k-1}),$$

where $r_k^* := \|x_k - x_h\|_*$. This contraction result readily implies the result of the theorem.

The complete proof is given in §A.5. □

4 Numerical Tests

4.1 Remarks on the implementation

Here, we remark on how preconditioning is implemented and on some further details of our implementation that slightly deviate from the theoretical formulations of Algorithms 1 and 3.

In all cases the underlying space is $X = \mathbb{R}^N$ for some $N \in \mathbb{N}$. The main deviation from Algorithms 1 and 3 is that we admit general Euclidean norms and inner products that may change from one step to another,

$$\|u\| = \sqrt{u^T M_k u}, \qquad \text{and} \qquad u \cdot v = u^T M_k v,$$

where M_k is symmetric and positive definite. That is, our implementation is a *variable metric variant*.

Let $E \in C^4(X) = C^4(\mathbb{R}^N)$, and let ∇' denote the standard gradient and $\otimes'$ the standard tensor product (i.e., the gradient and tensor products with respect to the ℓ^2-norm), then the gradient and tensor products in step k become

$$\nabla E(x) = M_k^{-1} \nabla' E(x), \quad \text{and} \quad (v \otimes v) \nabla E(x) = (v \otimes' v) \nabla' E(x).$$

The variable metric variant of Algorithm 1, augmented with a termination criterion, is given below. For the purposes of the numerical testing we call this the *simple dimer* method, it is effectively a forward Euler ODE integrator for the dimer dynamics. (Note also that here the rotation step is performed by a simple descent step followed by a projection, rather than a step on the manifold.)

Algorithm 1$^{\text{vm}}$:

1. **Input:** $x_0, v_0 \in X$, $h > 0$,$\alpha, \beta > 0$,$\mathrm{TOL}^x, \mathrm{TOL}^v > 0$; $k := 0$;
2. **While** $\|M_k^{-1/2} \nabla'_x \mathcal{E}_h(x_k, v_k)\|_{\ell^2} > \mathrm{TOL}^x$
   or $\|(M_k^{-1/2} - M_k^{1/2} v_k \otimes' v_k) h^{-2} \nabla'_v \mathcal{E}_h(x_k, v_k)\|_{\ell^2} > \mathrm{TOL}^v$ **do**
   %% Metric %%
3. Compute a spd matrix $M_k \in \mathbb{R}^{N \times N}$;
4. $v_k := v_k / \|M_k^{1/2} v_k\|$;
5. $v_{k+1} := v_k - \beta (M_k^{-1} - v_k \otimes v_k) h^{-2} \nabla'_v \mathcal{E}_h(x_k, v_k)$
6. $x_{k+1} := x_k - \alpha (M_k^{-1} - 2 v_k \otimes v_k) \nabla'_x \mathcal{E}_h(x_k, v_k)$.
7. $k := k + 1$

Remark 8. In our experiments we observe that the rotation residual decreases more quickly than the translation residual, hence the convergence criteria could be based on the translation residual only, without affecting the results. □

Analogous modifications are made to Algorithm 3. The auxiliary functional F_k now reads

$$\begin{aligned}
F_k(x) &= \mathcal{E}_h(x; v_k) - 2\big(v_k^T \nabla'_x \mathcal{E}_h(x_k, v_k)\big)\big(v_k^T M_k (x - x_k)\big) + \lambda_k \big(v_k^T M_k (x - x_k)\big)^2, \\
\lambda_k &= h^{-2} v_k^T \nabla'_v \mathcal{E}_h(x_k, v_k), \\
\nabla'_x \mathcal{E}_h(x, v) &= \tfrac{1}{2}\big(\nabla' E(x + hv) + \nabla' E(x - hv)\big), \\
\nabla'_v \mathcal{E}_h(x, v) &= \tfrac{h}{2}\big(\nabla' E(x + hv) - \nabla' E(x - hv)\big),
\end{aligned}$$

where we recall that ∇' denotes the standard gradient (i.e., the gradient with respect to the ℓ^2-norm).

Algorithm 3$^{\text{vm}}$:

1. **Input:** $x_0, v_0 \in X$, $h > 0$,$\mathrm{TOL}^x, \mathrm{TOL}^v > 0$; $k := 0$;
2. **While** $\|M_k^{-1/2} \nabla'_x \mathcal{E}_h(x_k, v_k)\|_{\ell^2} > \mathrm{TOL}^x$

%% Metric %%
3. Compute a spd matrix $M_k \in \mathbb{R}^{N\times N}$;
4. $v_k' := v_k / \|M_k^{1/2} v_{k-1}\|$;
%% Rotation %%
5. $v_{k+1} :=$ **Rotation (VM)**$[x_k, v_k', \beta, M_k]$
%% Translation %%
6. $p_M := -(M_k^{-1} - 2v_{k+1}\otimes v_{k+1})\nabla_x' \mathcal{E}_h(x_k; v_{k+1})$
7. $\alpha := \min(\alpha_{\max}, 2\alpha)$
8. **While** $(F_k(x_k + \alpha p_M) > F_k(x_k) - \Theta\alpha p_M^T M_k p_M)$
or $(\|M_k^{1/2}(M_k^{-1} - v_{k+1}\otimes v_{k+1})h^{-2}\nabla_v' \mathcal{E}_h(x_k + \alpha p_M; v_{k+1})\|_{\ell^2} >$
$\Psi\|M_k^{1/2}(M_k^{-1} - v_{k+1}\otimes v_{k+1})h^{-2}\nabla_v' \mathcal{E}_h(x_k; v_{k+1})\|_{\ell^2})$ **do**
9. $\alpha := \alpha/2$
10. $x_{k+1} := x_k + \alpha p_M$.
11. $k := k+1$

Rotation$^{\text{vm}}$:

1. **Input:** x, v, β, M_k
Parameters: $\text{TOL} = \max(\|M_k^{-1/2}\nabla_x' \mathcal{E}_h(x,v)\|_{\ell^2}, \text{TOL}^v)$, $\Theta \in (0,1)$, $\beta_{\max}$;
2. **While** $\|M_k^{1/2}(M_k^{-1} - v\otimes v)h^{-2}\nabla_v' \mathcal{E}_h(x; v)\| > \text{TOL}$ **do**
3. $s := -(M_k^{-1} - v\otimes v)h^{-2}\nabla_v' \mathcal{E}_h(x; v)$
4. $t := \|M_k^{1/2} s\|_{\ell^2}$; $\beta := \min(\beta_{\max}, 2\beta)$
5. $v_\beta := \cos(t\beta)v + \sin(t\beta)t^{-1}s$
6. **While** $\mathcal{E}_h(x, v_\beta) > \mathcal{E}_h(x,v) - \Theta\beta t^2$ **do**
7. $\beta := \beta/2$
8. $v := v_\beta$
9. **Output:** v, β

Remark 9. 1. In this formulation we compute only one metric update during each outer iterate. In particular, the metric remains fixed during the rotation step. Alternatively one could also allow metric updates during the rotation iterates.

2 An additional (optional) modification that can give significant performance gains is to employ a different heuristic for the initial guess of α in Step (7) of Algorithm 3$^{\text{vm}}$: With $p_{M,k} := -(M_k^{-1} - 2v_k\otimes v_k)\nabla_x' \mathcal{E}_h(x_k; v_k)$ and $p_{I,k} := -(I - 2v_k\otimes v_k)\nabla_x' \mathcal{E}_h(x_k, v_k)$ let, for $k \geq 2$, $\gamma_k := (p_{M,k-1} \cdot' p_{I,k-1})/(p_{M,k} \cdot' p_{I,k})$, then for $k \geq 2$ we replace Step (7) with

$$\alpha := \min\big(\text{avg}(\gamma_{\max(2,k-4)}, \ldots, \gamma_k), 2\alpha, \alpha_{\max}\big)$$

An analogous modification can be made for the rotation algorithm. □

In all numerical tests we use the following parameters: $h = 10^{-3}$, $\Theta = \sqrt{0.1}$, $\text{TOL}^x = 10^{-5}$, $\text{TOL}^v = 10^{-1}$, $\alpha_{\max} = 1$ and $\Psi = 100$. We briefly discuss these choices:

- h should be small enough such that the dimer saddle is sufficiently close to the true saddle (with respect to the length scales of the given problem), while large enough that numerical robustness does not become a problem for the rotation. In all our tests, $h = 10^{-3}$ was a good compromise.

- Θ should be sufficiently large (though, $\leq 1/2$) to ensure that the linesearch method finds steps which give a large decrease in dimer energy. It is often chosen much smaller than our choice of $\Theta = \sqrt{0.1}$ to immediately accept steps that make some progress. Our experience is that, with preconditioned search direction, our more stringent choice gives better performance.

- The choice of TOL^x simply controls the desired level of convergence to the dimer saddle.

- The parameter TOL^v should be chosen as weakly as possible such that either algorithm converges to the saddle. In Algorithm 3^{vm} rotations are performed such that the rotation residual is at least as good as the translation residual until it moves below this value. Subsequent translations may increase the rotation residual such that further applications of the rotation algorithm are needed. In practise this means that the rotation algorithm is performed at every iteration of Algorithm 3' for the first few steps, then only sporadically or not at all once the rotation residual reaches TOL^v. The use of this parameter then decreases the overall number of gradient evaluations needed to find the dimer saddle, by only performing the rotation as necessary.

- The maximum step $\alpha_{\max}$ should principally be chosen such that the dimer cannot translate into non-physical regimes for the given problem.

- The parameter Ψ should be chosen > 1 and restricts the translation step from moving the dimer to a point where it becomes too badly orientated. In our numerical tests this parameter is set sufficiently large that this termination criteria for the translation never occurs (the translation always terminates by finding a sufficient decrease in the auxiliary functional F_k).

Remark 10. 1. Since our convergence analysis is purely local it is relatively straightforward (though notationally tedious) to generalise it under mild assumptions on the metrics M_k. For example, in all our test cases, we have $M_k = M(x_k)$ where $M : X \to L(X)$ is continuous. In this case, the generalisation is straightforward.

2. We observe during numerical testing that the rotation component of the linesearch dimer is somewhat vulnerable to rounding error in the objective function E. As the dimer becomes increasingly well orientated, ∇E becomes almost orthogonal to the dimer orientation and any small rotation may result in a zero change (to numerical precision) in the dimer energy. In the numerical examples presented in this section, this never occurs since we use a relatively high value for TOL^v, that is the rotation is only ever weakly

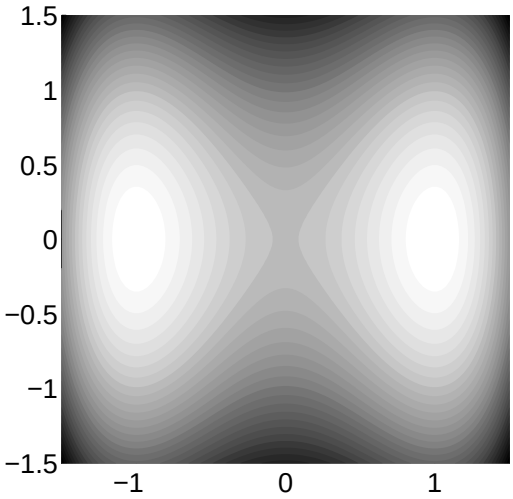

Figure 2: Energy function for Test 1 with 2 symmetric minima and a unique index-1 saddle

converged. In our examples this is sufficient for the the dimer to converge to the saddle. If a stronger level of converge were required, another technique should be used to improve the rotation residual further, such as changing to a gradient based method or simply making fixed steps. □

4.2 Test 1: A simple 2D example

Our first example is taken from [28]. We equip $X = \mathbb{R}^2$ with the standard Euclidean inner product. The energy function is given by $E(x, y) = (x^2 - 1)^2 + y^2$, which has two simple symmetric minima at $(\pm 1, 0)$ and a unique index-1 saddle at $(0, 0)$. The energy function is given graphically in Figure 2.

Figure 3 shows the x-residual $\|\nabla_x \mathcal{E}_h(x, v)\|$ plotted against the number of function evaluations and the number of iterations.

The performance of the linesearch dimer is compared with a simple dimer method with different step sizes. Evidently a good choice of step is important. If a poor choice is made the algorithm may perform poorly or diverge. The linesearch dimer method requires a certain amount of overhead versus a simple dimer with well chosen step sizes. We can see in Figure 3 that the linesearch dimer may find a solution in fewer dimer iterations than the best fixed step tested (indicating that it found better steps), but using more gradient evaluations.

4.3 Test 2: Vacancy Diffusion

Our second test case is a standard example from molecular physics. A single atom is removed from a 2D lattice and a neighbouring atom is moved partway into the gap. Atoms within a certain radius of the vacancy are allowed to move, while those beyond that radius

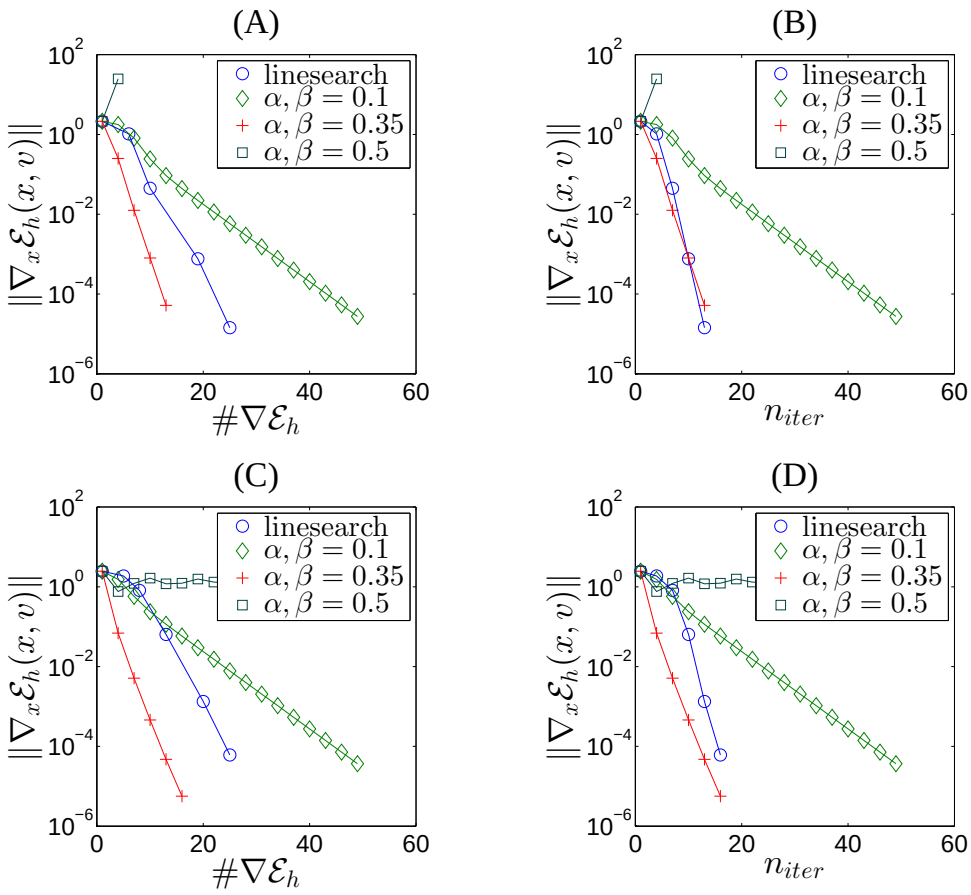

Figure 3: Convergence of the dimer to the saddle in a simple 2D example (Test 1).(A,B) The ℓ_2 norm of the x gradient versus the number of force evaluations and the number of dimer iteration where the initial dimer state is $x = [0.2, 1], v = [1, 1]$. In this case the choice $\alpha, \beta = 0.5$ diverges immediately. (C,D) The ℓ_2 norm of the x gradient versus the number of force evaluations and the number of dimer iterations where the initial dimer state is $x = [0.2, 1], v = [1, 1]$.

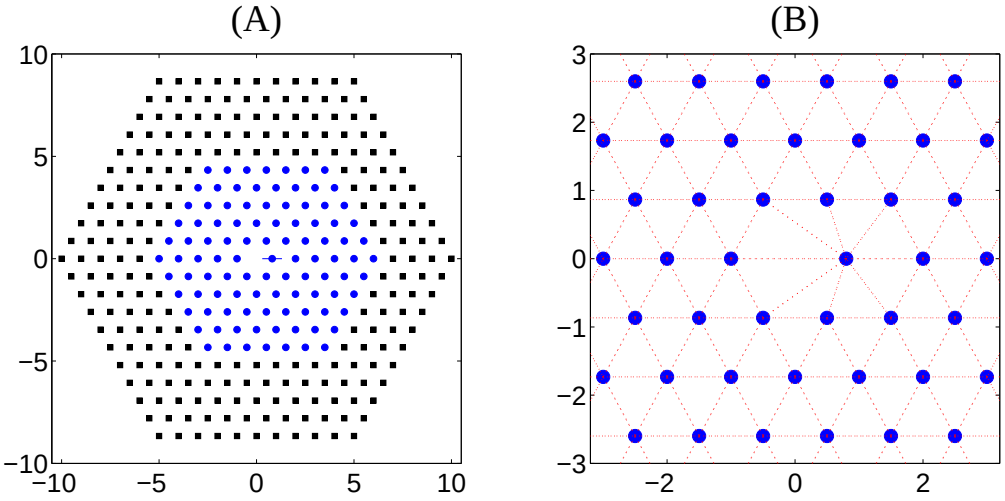

Figure 4: Initial configuration of the atoms in the vacancy diffusion problem (Test 2) . Black squares are fixed atoms while blue circles are atoms which move freely. (A) The initial dimer orientation is selected so that the translated atom has an orientation along the $y = 0$ direction, and is zero for all other atoms. (B) The Delaunay $\mathcal{T}_k$ triangulation used for the connectivity norm.

are fixed. This configuration is illustrated in Figure 4(A).

The energy function is given by the simple Morse potential,

$$E(\{x_i\}) = \sum_{i,j} V(\|x_i - x_j\|_2), \qquad V(r) = e^{-2a(r-1)} - 2e^{-a(r-1)}, \tag{15}$$

with stiffness parameter $a = 4$.

This test case demonstrates the importance of selecting the correct norm for high-dimensional problems. The experiment is run both using the generic ℓ_2 norm (no preconditioner), as well as a 'connectivity' norm. Such a norm can be defined based on the Delaunay triangulation of the atomistic positions (Figure 4(B))

$$\langle M_k u, u\rangle = \int |\nabla I_{\mathcal{T}_k} u|^2,$$

where $\mathcal{T}_k$ is the triangulation depicted in the figure and $I_{\mathcal{T}_k}$ the associated nodal interpolant.

Figure 5 demonstrates the convergence to the saddle with different numbers of free atoms nA (giving different dimensionality of the system) in the two norms for the linesearch dimer. We can also observe the benefit of the linesearch vs a simple dimer scheme when using the connectivity norm (Figure 6). The linesearch dimer selects very efficient stepsizes with no a-priori information, while the simple dimer method might exhibit either slow convergence, or no convergence, if the fixed steps are poorly chosen.

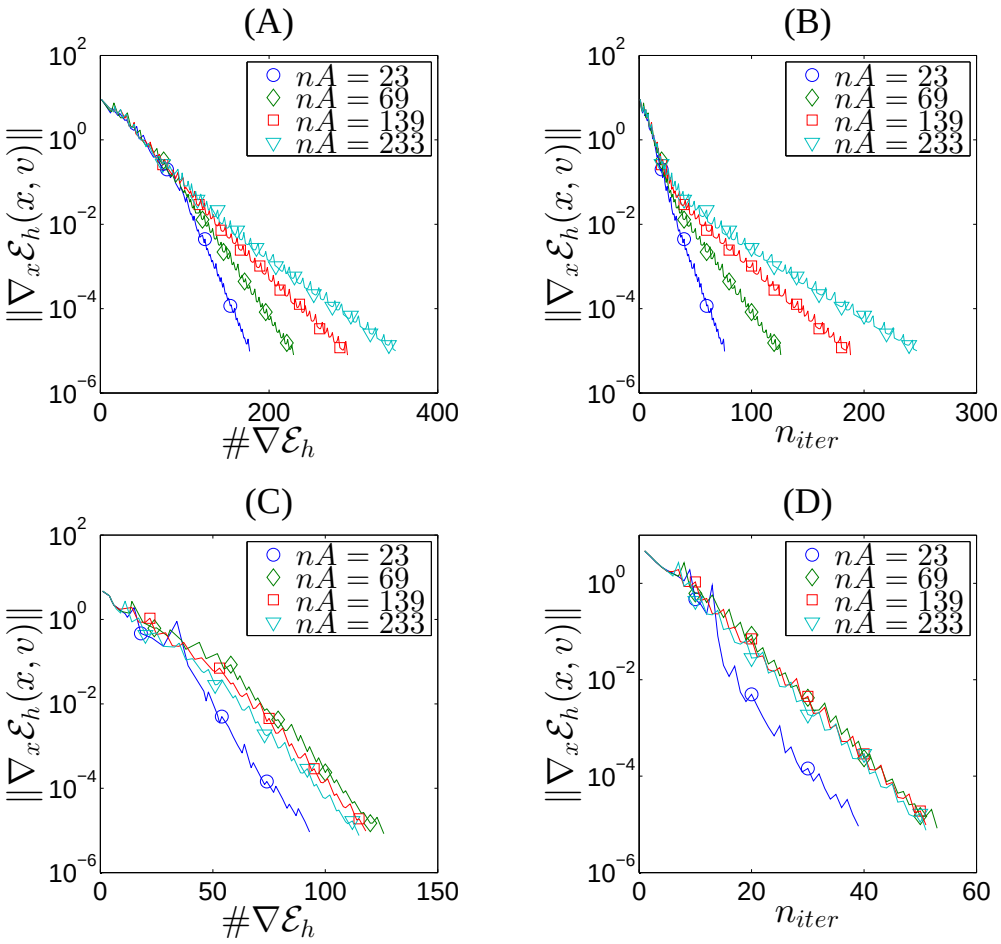

Figure 5: Convergence of the linesearch dimer to the saddle in the vacancy diffusion problem (Test 2) with (A),(C) the ℓ_2 norm and (B),(D) connectivity norm versus the number of force evaluations and dimer iterations for increasing numbers of free atoms.

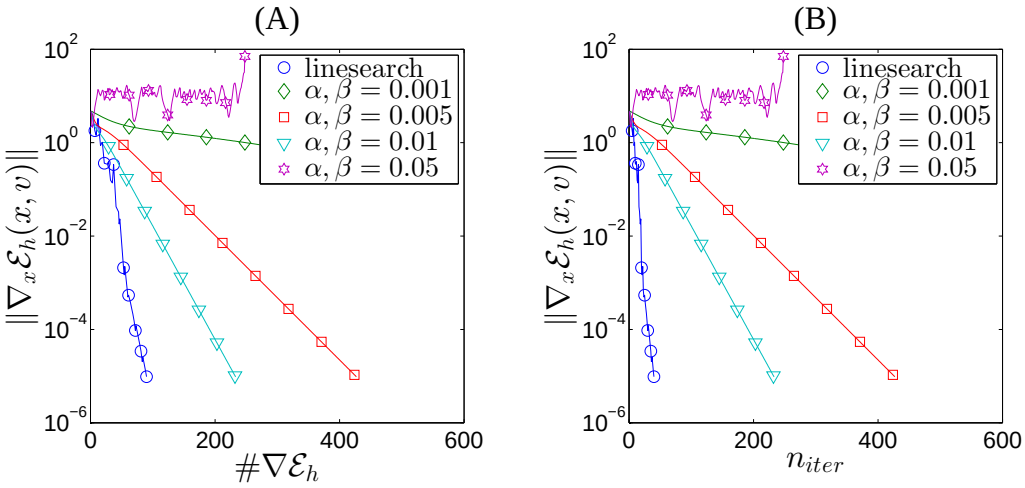

Figure 6: Convergence of the linesearch dimer vs the simple dimer method for Test 2 some choice of the of simple dimer step sizes with $nA = 69$ using the connectivity norm.

4.4 Test 3: A Phase Field Example

Our final example is based on a simple phase field model where the global energy is given by,

$$E(u) = \int_\Omega \frac{\epsilon}{2} |\nabla u|^2 + \frac{1}{2\epsilon}(u^2 - 1)^2. \tag{16}$$

In our test Ω is the unit square, and the boundary conditions are,

$$u(x) = \begin{cases} -1, & x_1 \in \{0, 1\} \\ 1, & x_2 \in \{0, 1\}. \end{cases} \tag{17}$$

There are 2 minima of such an energy, these are given in Figure 7(A),(B). The saddle between these two minima is given in Figure 7(C).

A possible choice for a preconditioner for this system is a stabilized Laplacian,

$$P = \epsilon \Delta + \frac{1}{\epsilon} I. \tag{18}$$

In order to compute either a minimum or a saddle point for such a system we triangulate the domain into a variable number of elements, thereby creating a discrete system of variable dimensionality. In our tests we take the initial dimer point as a small random perturbation of one of the local minima, and the initial dimer orientation is the metric inverted against a vector of ones.

In Figure 8 we demonstrate the necessity of using a preconditioner to solve this problem using the simple dimer method. When using the preconditioner (18), the algorithm performs well when the step size is chosen appropriately. We observe the expected behaviour, that there exists an optimal step size where convergence is fastest, and beyond

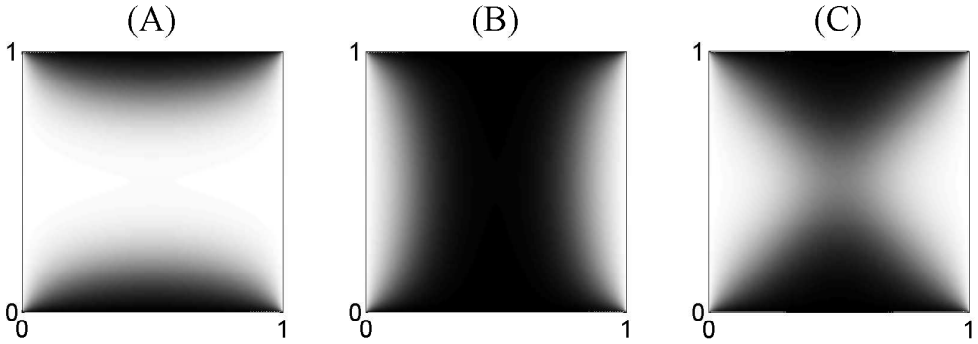

Figure 7: Minima (A,B) and saddle point (C) of the phase field problem (Test 3) with $\epsilon = 1/10$. The shading is linearly interpolated between white(-1) and black(1).

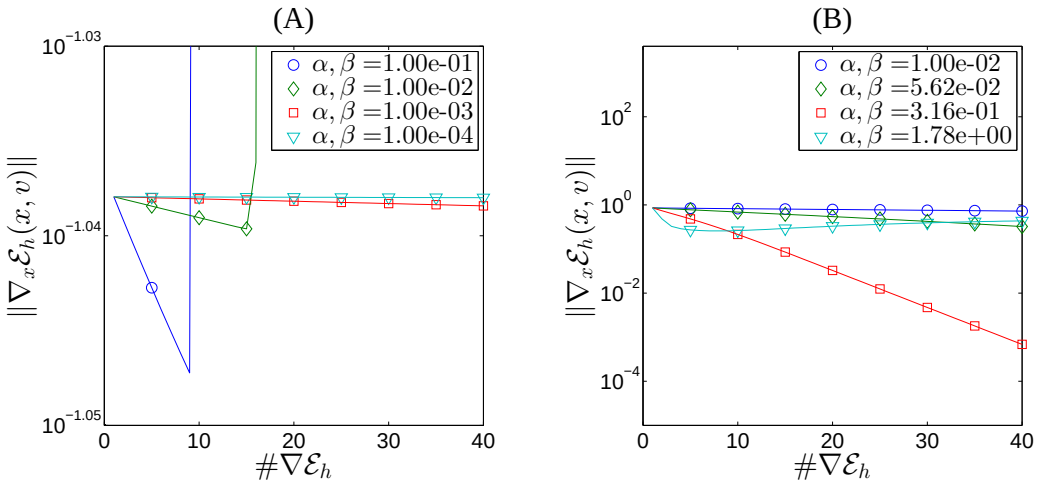

Figure 8: Convergence of the simple dimer to the saddle in the phase field problem (Test 3) with (A) the ℓ_2 metric and (B) the stabilized Laplacian metric where $\epsilon = 1/10$ for a triangulation with 3485 degrees of freedom.

that step size the dimer diverges. In fact we observe that the stabilized Laplacian metric is so effective, that the optimal step size seems very close to the unit step. If the ℓ_2 norm (identity preconditioner) is used then for all step sizes tested the dimer diverges, indicating that at best a very small step would need to be chosen for convergence.

In Figure 9 we demonstrate that the used of the scaled Laplacian metric for different system sizes. We observe that the use of this metric gives almost perfect scale invariance.

In Figure 10 we give the results of applying the simple and linesearch dimers with varying ϵ; the coarseness of the discretization in each experiment is chosen such that $\Delta x \approx \epsilon/5$. In some of these cases the linesearch dimer fails due to rounding error. Specifically, due to rounding error in the naive implementation of the energy function (simple summation over the elements), the translation step fails to find a sufficient decrease in the dimer energy, the step size selected shrinks to zero (to rounding error) and the method stagnates. In

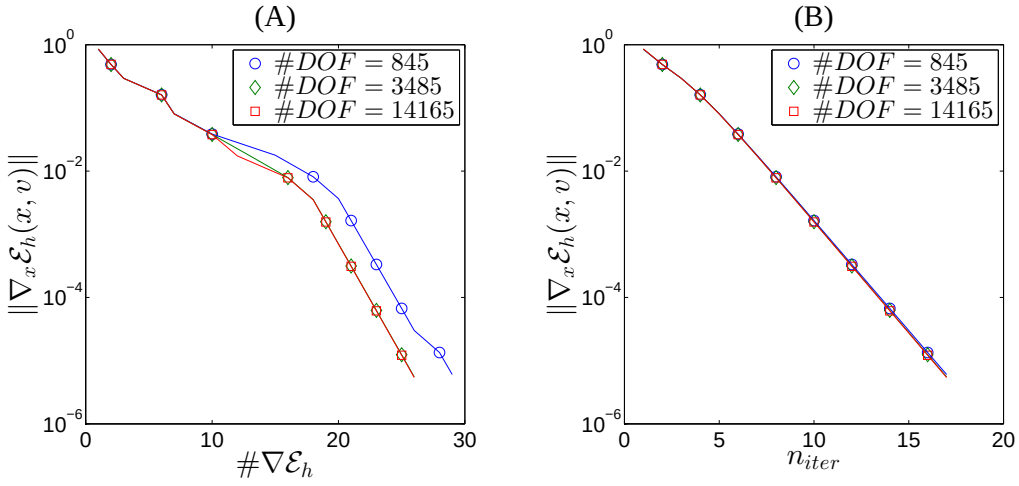

Figure 9: Convergence of the linesearch dimer to the saddle in the phase field problem (Test 3) with the stabilized Laplacian metric and triangulations of varying coarseness.

order to correct this a more robust method of evaluating the energy or a more advanced optimization algorithm should be implemented which can either choose better linesearch directions or more robustly deal with numerically zero energy changes.

We also observe, in the case $\epsilon = 1/30$ that the rate of convergence of even the simple dimer changes once the residual moves below a certain value. We are unable to give a satisfactory explanation for this effect, but speculate that the singularity in the boundary condition (which excludes admissible H^1-states) might be the case. (In particular, we observed that this behaviour is independent of the mesh coarseness and of the dimer length.)

5 Conclusions

We have described a dimer method for finding a saddle point in which the dimer length h is not required to shrink to zero, but which converges to a point that lies within $O(h^2)$ of a saddle. We have enhanced this algorithm with a lineasearch to improve its robustness, and use the observation that the dimer method may be formulated and applied in a general Hilbert space to allow preconditioning that improves the method's efficiency. The linesearch uses a local merit function. Unfortunately our particular merit function may not lead to global convergence of the iterates, and it is an open question as to whether there is another merit function that ensures global convergence. We have illustrated the positive effects of our algorithms on three realistic examples.

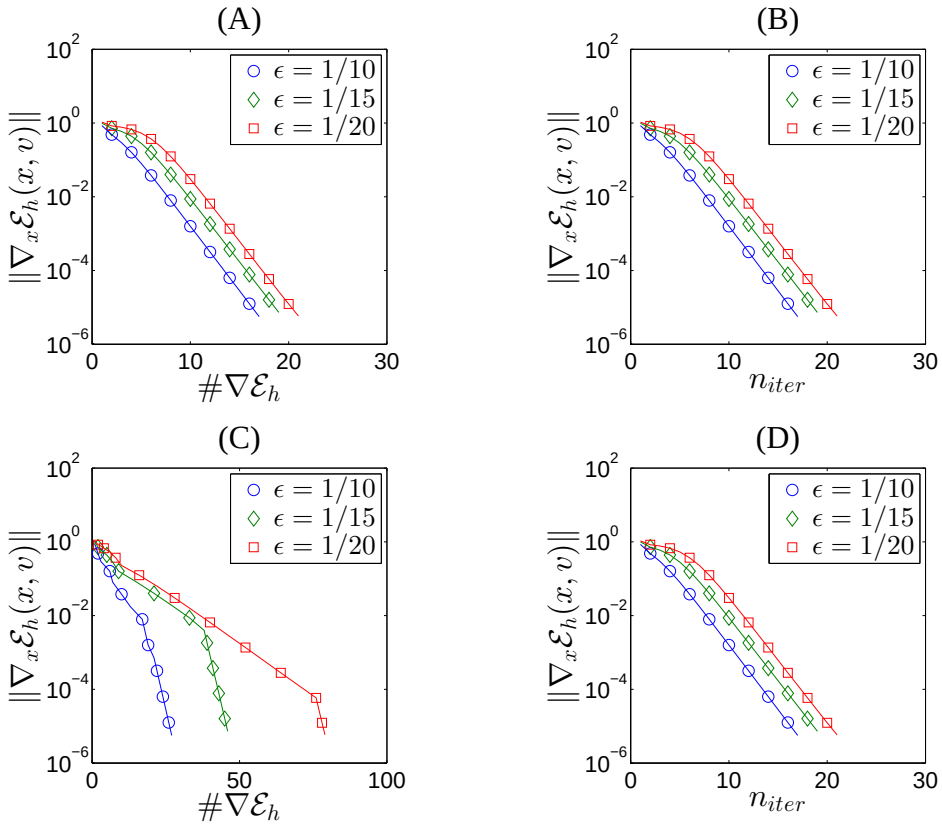

Figure 10: Convergence to the saddle in the phase field problem (Test 3) using the stabilized Laplacian metric with (A),(B) the simple dimer with unit step length and (C),(D) the linesearch dimer for a triangulation with 2405,9805,22205 degrees of freedom for the respective choices of ϵ.

A Proofs

A.1 Proof of Proposition 2

We prove the result using the inverse function theorem. We write (6) as $F(x_h, v_h, \lambda_h) = 0$ and show that $\|F(x_*, v_*, \lambda_*)\| \leq Ch^2$ and that $\nabla F(x_*, v_*, \lambda_*)$ is an isomorphism with bounds independent of h. The inverse function theorem then yields the stated result.

Residual estimate. Let the residual components be

$$\begin{aligned}
r_x &:= F_x(x_*, v_*, \lambda_*) = \tfrac{1}{2}\big(\nabla E(x_* + hv_*) + \nabla E(x_* - hv_*)\big),\\
r_v &:= F_v(x_*, v_*, \lambda_*) = \tfrac{1}{2h}(\nabla E(x_* + hv_*) - \nabla E(x_* - hv_*)) - \lambda_* v_*,\\
r_\lambda &:= F_\lambda(x_*, v_*, \lambda_*) = \tfrac{1}{2}(\|v_*\|^2 - 1).
\end{aligned}$$

Then,

$$\begin{aligned}
r_x &= \nabla E(x_*) + \tfrac{1}{2}\nabla^2 E(x_*)(hv_* - hv_*) + O(h^2) = O(h^2),\\
r_v &= \nabla^2 E(x_*)v_* - \lambda_* v_* + \tfrac{1}{h}\big(\nabla^3 E(x_*)[hv_* \otimes hv_* - hv_* \otimes hv_*] + O(h^2) = O(h^2),\\
r_\lambda &= 0.
\end{aligned}$$

Thus, $\|F(x_*, v_*, \lambda_*)\| \leq Ch^2$.

Stability. $\nabla F(x_*, v_*, \lambda_*)$ can be written in the form

$$\begin{aligned}
\nabla F(x_*, v_*, \lambda_*) &= \begin{bmatrix} \frac{\nabla^2 E(x_*+hv_*)+\nabla^2 E(x_*-hv_*)}{2} & h\frac{\nabla^2 E(x_*+hv_*)-\nabla^2 E(x_*-hv_*)}{2} & 0\\ \frac{\nabla^2 E(x_*+hv_*)-\nabla^2 E(x_*-hv_*)}{2h} & \frac{\nabla^2 E(x_*+hv_*)+\nabla^2 E(x_*-hv_*)}{2} - \lambda_* I & v_* \\ 0 & v_*^T & 0 \end{bmatrix}\\
&= \begin{bmatrix} \nabla^2 E(x_*) & 0 & 0\\ \nabla^3 E(x_*)\cdot v_* & \nabla^2 E(x_*) - \lambda_* I & v_* \\ 0 & v_*^T & 0 \end{bmatrix} + O(h^2) =: \mathbf{A} + O(h^2).
\end{aligned}$$

By assumption, $\nabla^2 E(x_*)$ is an isomorphism on X. Since, also by assumption, λ_* is a simple eigenvalue, the block

$$\begin{bmatrix} \nabla^2 E(x_*) - \lambda_* I & v_* \\ v_*^T & 0 \end{bmatrix} \tag{19}$$

is an isomorphism on $X \times \mathbb{R}$ as well. Thus, $\mathbf{A}$ is an isomorphism on $X \times X \times \mathbb{R}$ and consequently, for all h sufficiently small, $\nabla F(x_*, v_*, \lambda_*) = A + O(h^2)$ is also an isomorphism, with a uniform bound on its inverse.

Thus, the inverse function theorem shows that there exist a radius $r_0 > 0$ and a dimer length $h_0 > 0$, such that, for $h \leq h_0$, there exists a unique solution (x_h, v_h, λ_h) to (6) in a ball of radius r_0 about (x_*, v_*, λ_*), satisfying the estimate (7).

A.2 Proof of Theorem 3 (a)

Fix r and h_0 sufficiently small so that Theorem 2 applies. Let $e_k := x_k - x_h$, $f_k := v_k - v_h$ and $r_k := \sqrt{\|e_k\|^2 + \|f_k\|^2}$, so that trivially $\|e_k\| \leq r_k$ and $\|f_k\| \leq r_k$.

Lemma 11. *Let* $p := -(I - 2v_k \otimes v_k)\nabla_x \mathcal{E}(x_k, v_k)$ *and* $s := -(I - v_k \otimes v_k)H_h(x_k; v_k)$, *then, under the assumptions of Theorem 3,*

$$p = -Ae_k + O(r_k^2 + h^2 r_k), \qquad \text{and} \tag{20}$$
$$s = -Be_k - Cf_k + O(r_k^2 + h^2 r_k), \tag{21}$$

where the operators A *and* C *are defined in* (11) *and* B *is a bounded linear operator.*

Proof. To prove (20) we first note the following identities which are easy to establish:

$$\begin{aligned}
\nabla_x \mathcal{E}_h(x, v) &= \nabla E(x) + O(h^2), \\
\nabla_x \mathcal{E}_h(x_k, v_k) - \nabla_x \mathcal{E}_h(x_h, v_h) &= O(r_k), \\
v_k \otimes v_k - v_h \otimes v_h &= O(r_k), \\
\nabla_x^2 \mathcal{E}_h(x_h, v_h) &= \nabla^2 E(x_h) + O(h^2) = \nabla^2 E(x_*) + O(h^2) \\
\nabla_x \nabla_v \mathcal{E}_h(x_h, v_h) &= \tfrac{1}{2}\big(\nabla^2 E(x_h + hv_h) - \nabla^2 E(x_h - hv_h)\big) = O(h^2).
\end{aligned} \tag{22}$$

Using these identities, we can expand

$$\begin{aligned}
p &= -(I - 2v_k \otimes v_k)\big(\nabla_x \mathcal{E}_h(x_k, v_k) - \nabla_x \mathcal{E}_h(x_h, v_h)\big), \\
&= -(I - 2v_h \otimes v_h)\big(\nabla_x^2 \mathcal{E}_h(x_h, v_h)e_k + \nabla_x \nabla_v \mathcal{E}_h(x_h, v_h) f_k\big) + O(r_k^2) \\
&= -(I - 2v_h \otimes v_h)\nabla^2 E(x_h)e_k + O(r_k^2 + h^2 r_k) \\
&= -(I - 2v_* \otimes v_*)\nabla^2 E(x_*)e_k + O(r_k^2 + h^2 r_k) \\
&= -Ae_k + O(r_k^2 + h^2 r_k).
\end{aligned}$$

To prove (21), we first note that, with $\|v\| = 1$,

$$\begin{aligned}
H_h(x; v) &= \fint_{-1}^{1} \nabla^2 E(x + thv)\,\mathrm{d}t\, v = \nabla^2 E(x)v + O(h^2), \\
H_h(x_h; v_h) &= \nabla^2 E(x_h)v_h + O(h^2) = \nabla^2 E(x_*)v_* + O(h^2), \\
H_h(x_k; v_k) - H_h(x_h; v_h) &= \fint_{-1}^{1} \Big(\nabla^2 E(x_k + thv_k) - \nabla^2 E(x_h + thv_h)\Big)\,\mathrm{d}t\, v_k \\
&\qquad + \fint_{-1}^{1} \nabla^2 E(x_h + thv_h)\,\mathrm{d}t(v_k - v_h) \\
&= \fint_{-1}^{1} \Big(\nabla^3 E(x_h + thv_h)\big[(x_k - x_h) + th(v_k - v_h)\big]\,\mathrm{d}t v_h \\
&\qquad + \nabla^2 E(x_*)(v_k - v_h) + O(r_k^2 + h^2 r_k) \\
&= (\nabla^3 E(x_*)v_*)e_k + \nabla^2 E(x_*)f_k + O(h^2 r_k + r_k^2),
\end{aligned}$$

where we interpret $\nabla^3 E(x) \cdot v \in L(X)$ via the action $w \cdot ((\nabla^3 E(x) \cdot v)z) = \lim_{t\to 0} t^{-1} w \cdot$

$((\nabla^2 E(x+tv) - \nabla^2 E(x))z)$. Finally, we also have

$$\begin{aligned}(v_k \otimes v_k - v_h \otimes v_h)H_h(x_h; v_h) &= (v_k \otimes v_k - v_h \otimes v_h)\nabla^2 E(x_*)v_* + O(h^2 r_k)\\ &= \lambda_*(v_k \otimes v_k - v_h \otimes v_h)v_* + O(h^2 r_k)\\ &= \lambda_*(v_k \otimes v_k - v_h \otimes v_h)v_h + O(h^2 r_k)\\ &= \lambda_*(v_k - v_h) + \lambda_* v_k((v_k - v_h)\cdot v_h) + O(h^2 r_k)\\ &= \lambda_* f_k + O(r_k^2 + h^2 r_k).\end{aligned}$$

In the very last line we also used the fact that $v_k \cdot v_h - 1 = \frac{1}{2}\|v_k - v_h\|^2$.

Using these identities, we can compute

$$\begin{aligned}s &= -(I - v_k \otimes v_k)H_h(x_k; v_k)\\ &= (I - v_h \otimes v_h)H_h(x_h; v_h) - (I - v_k \otimes v_k)H_h(x_k; v_k)\\ &= -(I - v_k \otimes v_k)\big(H_h(x_k; v_k) - H_h(x_h; v_h)\big) + (v_k \otimes v_k - v_h \otimes v_h)H_h(x_h; v_h)\\ &= -(I - v_k \otimes v_k)\big((\nabla^3 E(x_*)v_*)e_k + \nabla^2 E(x_*)f_k\big) + O(h^2 r_k + r_k^2)\\ &\qquad + \lambda_* f_k + O(r_k^2 + h^2 r_k)\\ &=: -Be_k + \big[\lambda_* I - (I - v_* \otimes v_*)\nabla^2 E(x_*)\big]f_k + O(r_k^2 + h^2 r_k)\\ &= -Be_k - Cf_k + O(r_k^2 + h^2 r_k).\end{aligned}$$

□

From Lemma 11 it follows in particular that $s = O(r_k)$. Hence, Taylor expansions of sine and cosine in the identity

$$v_{k+1} = \cos\big(\|s\|\beta_k)v_k + \sin(\|s\|\beta_k)\tfrac{s}{\|s\|},$$

yield

$$f_{k+1} = f_k + \beta_k s + O(\beta_k^2 s_k^2)$$

Using Lemma 11, the identity $e_{k+1} = e_k + \alpha_k p$, and the fact that β_k is bounded, we therefore obtain identity (10) in the proof outline.

Upon defining

$$\mathbf{A}_k := \begin{pmatrix}\alpha_k A & 0\\ \beta_k B & \beta_k C\end{pmatrix} \quad \text{and} \quad \mathbf{e}_k = \begin{pmatrix}e_k\\ f_k\end{pmatrix}$$

(10) reads

$$\mathbf{e}_{k+1} = (I - \mathbf{A}_k)\mathbf{e}_k + \mathbf{t}_k, \tag{23}$$

where

$$\|\mathbf{t}_k\| \le C_{\mathrm{t}}(\alpha_k + \beta_k)(h^2 + r_k)r_k.$$

Due to the fact that A is symmetric and positive definite, it follows that, for $\bar{\alpha}, \bar{\beta}$ chosen sufficiently small and $\underline{\alpha} = \inf_k \alpha_k, \underline{\beta} = \inf_k \beta_k > 0$, the spectrum of $I - \mathbf{A}_k$ is real and belongs to $[0, 1-\epsilon]$ for some $\epsilon > 0$, that depends on $\underline{\alpha}, \underline{\beta}$. This will be crucial later in the proof.

Lemma 12. *Let* $\mathbf{P}_{n,k} := \prod_{i=n}^{k}(I - \mathbf{A}_i)$*, for* $0 \le n \le k$*, then there exist constants* $C_1 > 1, \mu \in (0,1)$ *such that*

$$\|\mathbf{P}_{n,k}\| \le C_1 \mu^{k-n+1}. \tag{24}$$

Proof. We assume, without loss of generality, that $n = 0$. First, we note that the diagonal blocks

$$A_k := [\mathbf{P}_{0,k}]_{xx} = \prod_{i=0}^{k}(I - \alpha_k A), \quad \text{and} \quad C_k := [\mathbf{P}_{0,k}]_{vv} = \prod_{i=0}^{k}(I - \beta_k C),$$

and it is easy to see that, for $\underline{\alpha} \le \alpha_i \le \bar{\alpha}, \underline{\beta} \le \beta_i \le \bar{\beta}$ chosen sufficiently small, that

$$\|A_k\| \le \bar{\mu}^k \qquad \text{and} \qquad \|C_k\| \le \bar{\mu}^k. \tag{25}$$

where $\bar{\mu} := \max\big(1 - \underline{\alpha}\inf\sigma(A), 1 - \underline{\beta}\inf\sigma(C)\big) \in (0,1)$.

Since the off-diagonal block $[\mathbf{P}_{0,k}]_{xv} = 0$, it remains to estimate the off-diagonal block $B_k := [\mathbf{P}_{0,k}]_{vx}$. We use induction over k. Let $C_* := \bar{\mu}^{-1}\underline{\beta}\|B\|$ and suppose that

$$\|B_k\| \le C_* k \bar{\mu}^k. \tag{26}$$

Then, using

$$B_{k+1} = -\beta_{k+1} B A_k + (I - \beta_{k+1} C) B_k,$$

as well as $\|I - \beta_{k+1}C\| \le \bar{\mu}$ we can estimate

$$\|B_{k+1}\| \le \underline{\beta}\|B\|\bar{\mu}^k + \bar{\mu} C_* k \bar{\mu}^k = C_*(k+1)\bar{\mu}^{k+1},$$

which establishes the induction since the result is true by definition of C_* when $k = 0$.

Now pick $\tau > 1$ so that $\mu := \tau\bar{\mu} < 1$. Then (25) give that $\|A_k\| \le \mu^k$ and $\|C_k\| \le \mu^k$, while it follows from (26) and by maximizing $x\tau^{-x}$ that

$$\|B_k\| \le C_*(k/\tau^k)\mu^k \le (C_*/\tau \ln \tau)\mu^k.$$

The result now follows from the inequality

$$\|\mathbf{P}_{0,k}\| \le \big(\|[\mathbf{P}_{0,k}]_{xx}\|^2 + \|[\mathbf{P}_{0,k}]_{vx}\|^2 + \|[\mathbf{P}_{0,k}]_{vv}\|^2\big)^{1/2},$$

and by defining $C_1 := \mu^{-1}\sqrt{2 + (C_*/\tau\ln\tau)^2}$.

□

It is straightforward to prove that

$$\mathbf{e}_{k+1} = \mathbf{P}_{0,k}\mathbf{e}_0 + \mathbf{t}_k + \mathbf{P}_{k,k}\mathbf{t}_{k-1} + \mathbf{P}_{k-1,k}\mathbf{t}_{k-2} + \cdots + \mathbf{P}_{1,k}\mathbf{t}_0,$$

which implies

$$\|\mathbf{e}_{k+1}\| \le C_1\mu^{k+1}\|\mathbf{e}_0\| + \sum_{i=0}^{k} C_1\mu^i\|\mathbf{t}_{k-i}\|,$$

that is,

$$r_{k+1} \le C_2\Big(\mu^{k+1} r_0 + \sum_{i=0}^{k} \mu^{i+1}(r_{k-i} + h^2) r_{k-i}\Big), \tag{27}$$

for some $C_2 \ge C_1$.

We make another induction hypothesis that,

$$r_i \le C_3 \gamma^i r_0, \tag{28}$$

where $\gamma \in (\mu, 1)$ and $C_3 > C_2$ are arbitrary. The statement (28) is clearly true for $i = 0$. Assume now that it holds for $i = 0, \dots, k$, then (27), and using $\mu/\gamma < 1$ yields

$$\begin{aligned} r_{k+1} &\le C_2\gamma^{k+1} r_0 \Big(\big(\tfrac{\mu}{\gamma}\big)^{k+1} + \sum_{i=0}^{k} \big(\tfrac{\mu}{\gamma}\big)^{i+1} \Big(C_3 \gamma^{k-i} r_0 + h^2 \Big) C_3 \Big) \\ &\le C_2 \gamma^{k+1} r_0 \Big(1 + \frac{C_3^2 r_0 + C_3 h^2}{1 - \mu/\gamma} \Big). \end{aligned}$$

Since $C_3 > C_2$, upon choosing r_0, h sufficiently small, we can achieve that

$$C_2\Big(1 + \frac{C_3^2 r_0 + C_3 h^2}{1 - \mu/\gamma} \Big) \le C_3,$$

hence (28) holds also for $i = k + 1$. This completes the proof of (28) and hence of Theorem 3 (a).

A.3 Proof of Theorem 3 (b)

We begin with a basic auxiliary result.

Lemma 13. *Let (x_*, v_*, λ_*) be an index-1 saddle and $\mu_* := \inf_{\|w\|=1, w \perp v_*} (\nabla^2 E(x_*) w) \cdot w > 0$. Then, there exists $r > 0$ and $h_0 > 0$ (chosen independently of one another) such that the following hold:*

(i) If $x \in B_r(x_*)$ then $\nabla^2 E(x)$ has index-1 saddle structure and, if (λ, v) is the smallest eigenpair of $\nabla^2 E(x)$, then $\lambda \le \lambda_*/2$ and $(\nabla^2 E(x) w) \cdot w \ge \mu_*/2 \|w\|^2$ for $w \perp v$.

(ii) $V(x)$ is well-defined for all $x \in B_r(x_*)$ and $h \in (0, h_0]$, and $x \mapsto V(x) \in C^1(B_r(x_*))$.

(iii) $E_h \in C^4(B_r(x_*))$ with

$$\begin{aligned} \nabla E_h(x) = \nabla_x \mathcal{E}_h(x, V(x)) = \tfrac{1}{2}\big(\nabla E(x + hV(x)) + \nabla E(x - hV(x))\big), \quad \text{and} \\ \|\nabla^2 E_h(x) - \nabla^2 E(x)\| \le C_0 h^2 \end{aligned}$$

for $x \in B_r(x_*)$, where C_0 is independent of x, h.

(iv) Let $x \in B_r(x_*)$ and let (λ, v) be the minimal eigenpair of $\nabla^2 E_h(x)$, then $\|v - V(x)\| \le Ch^2$, where C is independent of x, h.

Proof. For r sufficiently small, the statement (i) is an obvious consequence of x_* being an index-1 saddle and $\nabla^2 E$ locally Lipschitz continuous (which follows since $E \in C^4(X)$).

The statement (ii) is proven similarly as Proposition 2, provided h_0 is chosen sufficiently small (depending on λ_*, μ_* and on derivatives of E in $B_{2r}(x_*)$). The C^1-dependence of $V(x)$ on x is a consequence of the implicit function theorem.

The statement (iii) follows from an elementary Taylor expansion.

Finally, (iv) follows again from (iii) and an argument analogous to Proposition 2. □

To complete the proof of Theorem 3(b) we first note that, according to Lemma 13(ii), Step (2) of Algorithm 2 is indeed well-defined, provided that we can ensure that the iterates never leave a neighbourhood of x_h and hence of x_*. This will be established.

Fix r, h_0 sufficiently small so that Theorem 2 and Lemma 13 apply. Let $e_k := x_k - x_h$ and $r_k := \|e_k\|$. Let $s := -(I - 2v_k \otimes v_k)\nabla E_h(x_k)$ be the search direction and $\alpha_k > 0$ the step size, then

$$e_{k+1} = e_k + \alpha_k s$$

Applying Lemma 13(iii) we can expand

$$\nabla E_h(x_k) = \nabla E_h(x_k) - \nabla E_h(x_h) = \nabla^2 \mathcal{E}_h(x_h) e_k + O(r_k^2) = \nabla^2 E(x_*) e_k + O(h^2 r_k + r_k^2).$$

Arguing similarly as in the proof of part (a),

$$\begin{aligned} e_{k+1} &= e_k - \alpha_k (I - 2v_k \otimes v_k)\nabla^2 E(x_*) e_k + O(h^2 r_k) \\ &= e_k - \alpha_k (I - 2v_* \otimes v_*)\nabla^2 E(x_*) e_k + O\big(h^2 r_k + r_k^2\big) \\ &= (I - \alpha_k A) e_k + O\big(h^2 r_k + r_k^2\big). \end{aligned}$$

For $\bar{\alpha}$ sufficiently small it is straightforward to see that $\|I - \alpha_k A\| \le 1 - \alpha_k \epsilon \le 1 - \underline{\alpha}\epsilon =: \gamma$, where $\epsilon > 0$ and $\gamma \in (0, 1)$, and we therefore obtain

$$r_{k+1} \le (\gamma + C_1 h^2 + C_2 r_k) r_k.$$

Clearly, for h_0 and r_0 chosen sufficiently small we obtain a contraction, that is, $r_{k+1} \le \gamma' r_k$ for some $\gamma' \in (\gamma, 1)$.

This completes the proof of Theorem 3(b).

A.4 Contraction of steepest descent with linesearch

In the section following this one, we will use statements about the steepest descent method with backtracking that we suspect must be well known. Since we have been unable to find precisely the versions we require, we give both below, the latter with a full proof.

Lemma 14. *Let X be a Hilbert space, $F \in C^3(X)$, and $x_* \in X$ with $\nabla F(x_*) = 0$ and $\nabla^2 F(x_*)$ positive definite, i.e., $u \cdot (\nabla^2 F(x_*) u) \ge \mu \|u\|^2$ for $\mu > 0$. Let $\|u\|_*^2 := u \cdot (\nabla^2 F(x_*) u)$. Further, let $\bar{\alpha} > \underline{\alpha} > 0$, $\Theta \in (0, 1)$.*

Then, there exists $r > 0$ and $\gamma \in (0,1)$, depending only on $\underline{\alpha}, \bar{\alpha}, \mu, \|\nabla^j F(x)\|$ for $x \in B_1(x_*)$, such that, for all $\alpha \in [\underline{\alpha}, \bar{\alpha}]$ and for all $x \in B_r(x_*)$ satisfying the Armijo condition

$$F(x - \alpha \nabla F(x)) \leq F(x) - \Theta \alpha \|\nabla F(x)\|^2,$$

we have

$$\big\|[x - \alpha \nabla F(x)] - x_*\big\|_* \leq \gamma \|x - x_*\|_*.$$

Proof. The proof is a simplified version of the proof of Lemma 15 below. □

We now generalize the foregoing result to steepest descent on the unit sphere. Convergence results for many methods on manifolds are given by [1, Chap.4]. See specifically [1, Thm.4.5.6] and [2].

Lemma 15. *Let X be a Hilbert space, $S_X := \{u \in X \,|\, \|u\| = 1\}$, $P_v := v \otimes v$ and $P_v' := I - P_v$ for $v \in S_X$. Let $F \in C^3(X)$,*

$$g(v) := P_v' \nabla F(v) \qquad \text{and} \qquad H(v) := P_v' \nabla^2 F(v) P_v' - \big(\nabla F(v) \cdot v\big) I.$$

We assume that there exists $v_* \in S_X$ and $\mu > 0$ such that

$$g(v_*) = 0 \qquad \text{and} \qquad u \cdot \big(H(v_*) u\big) \geq \mu \|u\|^2 \quad \forall u \in X. \tag{29}$$

Let $\|u\|_* := \sqrt{u \cdot (H(v_*)u)}$.

Let $\bar{\alpha} > 0$, $\Theta \in (0,1)$, and for $v \in S_X$ and $\alpha \in \mathbb{R}$, denote

$$v_\alpha := \cos\big(\alpha \|g(v)\|\big) v - \sin\big(\alpha \|g(v)\|\big) \frac{g(v)}{\|g(v)\|}.$$

Then, there exists $r > 0$ such that, for all $v \in B_r(v_*) \cap S_X$ and $\alpha \in (0, \bar{\alpha}]$ satisfying the Armijo condition

$$F(v_\alpha) \leq F(v) - \Theta \alpha \|g(v)\|^2,$$

there exists a constant $\gamma(\alpha) \in [0,1)$ such that

$$\big\|v_\alpha - v_*\big\|_* \leq \gamma(\alpha) \|v - v_*\|_*.$$

The contraction factor $\gamma(\alpha)$ depends on α, μ and on $\|\nabla^j F(x)\|, x \in B_1(v_*)$. Moreover, for any $\underline{\alpha} \in (0, \bar{\alpha}]$, $\sup_{\alpha \in [\underline{\alpha}, \bar{\alpha}]} \gamma(\alpha) < 1$.

Proof. We first note that $\|\cdot\|_*$ is an equivalent norm, that is, there exists a constant $C_* = \|H(v_*)\|$ such that

$$\sqrt{\mu} \|u - u'\| \leq \|u - u'\|_* \leq C_* \|u - u'\| \qquad \forall u, u' \in X. \tag{30}$$

Step 1: Expansions. There exists a constant C_L such that, for all $v, w \in S_X$,

$$\big\|g(v) - g(w)\big\| \le C_L\|v - w\|, \quad \text{and} \tag{31}$$
$$\big\|\nabla^2 F(v) - \nabla^2 F(w)\big\| \le C_L\|v - w\|. \tag{32}$$

since $F \in C^3(X)$ and S_X is bounded. For $v \in S_X$ the identity

$$v_* \cdot (v - v_*) = -\tfrac{1}{2}\|v - v_*\|^2 \tag{33}$$

and $g(v_*) = 0$ yields

$$\nabla F(v_*) \cdot (v - v_*) = \nabla F(v_*) \cdot \big((v_* \otimes v_*)(v - v_*)\big) = \big(-\tfrac{1}{2}\nabla F(v_*) \cdot v_*\big)\|v - v_*\|^2, \tag{34}$$

and therefore,

$$\begin{aligned} F(v) - F(v_*) &= \nabla F(v_*) \cdot (v - v_*) + \left(\int_0^1 (1-t)\nabla^2 F((1-t)v_* + tv)\,\mathrm{d}t(v - v_*) \right) \cdot (v - v_*) \\ &= \tfrac{1}{2}(v - v_*) \cdot \big(\bar{H}_v(v - v_*)\big) + \tfrac{1}{2}(v - v_*) \cdot \big(\big[\nabla^2 F(v_*) - \big(\nabla F(v_*) \cdot v_*\big)I\big](v - v_*)\big), \end{aligned} \tag{35}$$

where

$$\bar{H}_v := 2\int_0^1 (1-t)[\nabla^2 F((1-t)v_* + tv) - \nabla^2 F(v_*)]\,\mathrm{d}t.$$

But

$$(v - v_*) \cdot \Big(\nabla^2 F(v_*)(v - v_*)\Big) = (v - v_*) \cdot \Big(P'_{v_*}\nabla^2 F(v_*)P'_{v_*}(v - v_*)\Big) + O(\|v - v_*\|^3).$$

since $(v_* \otimes v_*)(v - v_*) = O(\|v - v_*\|^2)$, and thus we obtain from (32) and (35) that

$$\tfrac{1}{2}\|v - v_*\|_*^2 - C_1\|v - v_*\|^3 \le F(v) - F(v_*) \le \tfrac{1}{2}\|v - v_*\|_*^2 + C_1\|v - v_*\|^3, \tag{36}$$

for some constant C_1 that depends on C_L.

Step 2: Bound on descent step. The Lipschitz bound (31) implies that, for all $v \in S_X$,

$$\begin{aligned} \|v_\alpha - v_*\| &\le \|v - v_*\| + \big|1 - \cos(\alpha\|g(v)\|)\big| + \big|\sin(\alpha\|g(v)\|)\big| \\ &\le \|v - v_*\| + \tfrac{1}{2}(\alpha\|g(v)\|)^2 + \alpha\|g(v)\| \\ &\le \|v - v_*\| + \tfrac{1}{2}\alpha^2 C_L^2\|v - v_*\|^2 + \alpha C_L\|v - v_*\| \\ &\le \big(1 + \alpha^2 C_L^2 + \alpha C_L\big)\|v - v_*\| \\ &=: c_3(\alpha)\|v - v_*\|, \end{aligned}$$

as $|1 - \cos\theta| \le \frac{1}{2}\theta^2$ and $|\sin\theta| \le \theta$ for $\theta \ge 0$, and $\|v - v_*\| \le 2$. In particular, for $r > 0$

$$\|v_\alpha - v_*\| \le c_3(\bar{\alpha})r \qquad \forall v \in B_r(v_*) \cap S_X. \tag{37}$$

Step 3. Bound on gradient. To obtain an error estimate from the Armijo condition, we must bound $\|g(v)\|^2$ below. We write $v_t := (1-t)v_* + tv$, then

$$\begin{aligned}
g(v) &= g(v) - g(v_*) \\
&= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\Big((I - v_t \otimes v_t)\nabla F(v_t)\Big)\,\mathrm{d}t \qquad (38)\\
&= \int_0^1 \Big((I - v_t \otimes v_t)\nabla^2 F(v_t)(v - v_*) - \big((v - v_*)\otimes v_t + v_t \otimes (v - v_*)\big)\nabla F(v_t)\Big)\,\mathrm{d}t \\
&= (I - v_* \otimes v_*)\nabla^2 F(v_*)(v - v_*) \qquad (39)\\
&\quad - \big((v - v_*)\otimes v_* + v_* \otimes (v - v_*)\big)\nabla F(v_*) + O(\|v - v_*\|^2) \\
&= H(v_*)(v - v_*) + (I - v_* \otimes v_*)\nabla^2 F(v_*)(v_* \otimes v_*)(v - v_*) \\
&\quad - \big(\nabla F(v_*)\cdot(v - v_*)\big)v_* + O(\|v - v_*\|^2) \\
&= H(v_*)(v - v_*) + O(\|v - v_*\|^2), \qquad (40)
\end{aligned}$$

where we used (33) and (34) in the last step.

Thus, for some constant C_2 that depends only on C_L, and for $v \in B_r(v_*) \cap S_X$, with $r \le r_1$ and r_1 chosen sufficiently small, we obtain

$$\begin{aligned}
\|g(v)\|^2 &\ge \|H(v_*)(v - v_*)\|^2 - C_2\|v - v_*\|^3 \\
&\ge \mu\|H(v_*)^{1/2}(v - v_*)\|^2 - C_2\|v - v_*\|^3 \\
&\ge \big(\mu - C_2\mu^{-1}r\big)\|v - v_*\|_*^2 \\
&\ge \tfrac{\mu}{2}\|v - v_*\|_*^2 \qquad (41)
\end{aligned}$$

using (29) and (30).

Step 4. Short steps. For α sufficiently small, the Armijo condition is in fact not needed, and we can proceed without it. From the definition of v_α and Taylor's theorem we obtain, for $\alpha \le \bar{\alpha}$

$$v_\alpha - v = \alpha g(v) + O(\alpha^2\|g(v)\|^2)$$

and hence using (40)

$$v_\alpha - v_* = \big[I - \alpha H(v_*)\big](v - v_*) + O(\alpha\|v - v_*\|^2)$$

Taking the inner product with $H(v_*)(v_\alpha - v_*)$, there exists a constant c_4 that depends only on the derivatives F in $B_1(v_*)$ such that

$$\|v_\alpha - v_*\|_*^2 \le (v - v_*)\cdot\Big(H(v_*)\big[I - \alpha H(v_*)\big]\Big)(v - v_*) + c_4 r\alpha\|v - v_*\|_*^2.$$

The eigenvalues of $H(v_*)\big[I - \alpha H(v_*)\big]\psi = \tau H(v_*)\psi$ are precisely $\tau = 1 - \alpha\lambda$ for $\lambda \in \sigma(H(v_*))$. Let $\hat{\alpha} > 0$ such that $\tau \in [0, 1)$ for all $\alpha \le \hat{\alpha}$. Then, the largest eigenvalue is given by $1 - \alpha\mu$ and we obtain that, for $\alpha \le \hat{\alpha}$,

$$\|v_\alpha - v_*\|_*^2 \le (1 - \alpha\mu + c_4\alpha r)\|v - v_*\|_*^2.$$

Choosing $r \le r_2 \le r_1$ sufficiently small, with the new restrictions depending only on μ and c_4, and using the bound $\sqrt{1-\theta} \le 1 - \frac{1}{2}\theta$ for $\theta \in [0,1]$, we obtain that

$$\|v_\alpha - v_*\|_* \le (1 - \alpha\tfrac{\mu}{4})\|v - v_*\|_*.$$

This completes the proof of the Lemma, for the case $\alpha \le \hat{\alpha}$.

Step 4. Long steps. Let $\alpha \in [\hat{\alpha}, \bar{\alpha}]$, $r \le r_1$, $v \in B_r(v_*) \cap S_X$, $c_3 \equiv c_3(\bar{\alpha})$, and v_α satisfying the Armijo condition, then (30), (36), (37) and (41) imply

$$\begin{aligned}\big(\tfrac{1}{2} - C_1\mu^{-1}c_3 r\big)\|v_\alpha - v_*\|_*^2 &\le F(v_\alpha) - F(v_*) \\ &\le F(v) - F(v_*) - \Theta\alpha\|g(v)\|^2 \\ &\le \big(\tfrac{1}{2} + C_1\mu^{-1}r - \Theta\hat{\alpha}\tfrac{\mu}{2}\big)\|v - v_*\|_*^2,\end{aligned}$$

that is,

$$\|v_\alpha - v_*\|_* \le \left(\frac{1 + 2C_1\mu^{-1}r - \Theta\hat{\alpha}\mu}{1 - 2C_1\mu^{-1}c_3 r}\right)^{1/2}\|v - v_*\|_*.$$

Thus, choosing $r \le r_1$, sufficiently small, we obtain again the desired contraction. □

A.5 Proof of Theorem 7

Throughout this proof, we fix an index-1 saddle (x_*, v_*, λ_*), and assume that h_0 is small enough so that Proposition 2 ensures the existence of a dimer saddle (x_h, v_h, λ_h) in an $O(h^2)$ neighbourhood of (x_*, v_*, λ_*).

The first step is an error bound on $v_k - v_h$ in terms of $x_k - x_h$ and the residual of v_k.

Lemma 16. *There exist $r, h_0, C_1 > 0$ such that, for $h \in (0, h_0]$, $x \in B_r(x_*)$ and $v \in B_r(v_*)$ with $\|v\| = 1$, we have*

$$\|v - v_h\| \le \tfrac{1}{2}C_1\big(\|x - x_h\| + \big\|(I - v\otimes v)H_h(x;v)\big\|\big).$$

Proof. Let $\lambda := H_h(x;v)\cdot v$, then

$$\begin{aligned} H_h(x_h;v) &= \lambda v + s, \\ \tfrac{1}{2}\|v\|^2 &= \tfrac{1}{2}, \end{aligned} \tag{42}$$

where

$$s = \big(H_h(x_h;v) - H_h(x;v)\big) + (I - v\otimes v)H_h(x;v).$$

Since v_h solves (42) with $s = 0$, and since

$$\|s\| \le C_2\big(\|x - x_h\| + \|(I - v\otimes v)H_h(x;v)\|\big),$$

the stated result follows from the Lipschitz continuity of $H_h(\cdot;v)$ and an application of the inverse function theorem, in a similar spirit as the proof in §A.1. □

Next, we present a result ensuring that the rotation step of Algorithm 3 not only terminates but also produces a new dimer orientation v_k which remains in a small neighbourhood of the "exact" orientation v_h.

Lemma 17. *There exist $r, h_0, C_2 > 0, C_3 \geq 1$ such that, if $h \in (0, h_0]$, $x_k \in B_r(x_*)$, $v_{k-1} \in B_{C_3 r}(v_*)$, $\|v_{k-1}\| = 1$, then Step (3) of Algorithm 3 terminates with outputs $v_k \in B_{C_3 r}(v_*)$, $\|v_k\| = 1$, $\beta_k > 0$, satisfing*

$$\|v_k - v_h\| \leq C_2\big(\|x_k - x_h\| + h^2\|v_{k-1} - v_h\|\big). \tag{43}$$

Proof. Let $G(v) := h^{-2}(\mathcal{E}_h(x_k; v) - \mathcal{E}_h(x_k; V(x_k)))$, then each step of the Rotation Algorithm is a steepest descent step of G on the manifold $S_X := \{\|v\| = 1\}$. We need to ensure that these iterations do not "escape" from the minimiser.

Lemma 15 (with $F(v) = G(v)$ and $v_* \equiv V(x_k)$) implies that each such step is a contraction towards $V(x_k)$ with respect to the norm $\|\cdot\|_H$ induced by the operator

$$H := (I - V \otimes V)\nabla^2 G(V)(I - V \otimes V) - (\nabla G(V) \cdot V)I,$$

where $V \equiv V(x_k)$; provided that r is sufficiently small and H is positive definite.

To see that the latter is indeed true, we recall from (2) and (3) that

$$\nabla G(V(x_k)) = \nabla^2 E(x_k)V(x_k) + O(h^2) \quad \text{and} \quad \nabla^2 G(V(x_k)) = \nabla^2 E(x_k) + O(h^2)$$

and from Proposition 2 and Lemma 13 that

$$V(x_k) = v_* + O(h^2 + r), \tag{44}$$

and hence,

$$\begin{aligned} H &= (I - v_* \otimes v_*)\nabla^2 E(x_*)(I - v_* \otimes v_*) - \big((\nabla^2 E(x_*)v_*) \cdot v_*\big)I + O(h^2 + r) \\ &= (I - v_* \otimes v_*)\nabla^2 E(x_*) - \lambda_* I + O(h^2 + r). \end{aligned}$$

Since (x_*, v_*, λ_*) is an index-1 saddle, $(I - v_* \otimes v_*)\nabla^2 E(x_*)$ is positive definite in $\{v_*\}^\perp$, and $\lambda_* < 0$. Thus, for h, r sufficiently small, H is positive definite as required.

From Lemma 15, it follows that all iterates $v_k^{(j)}$ of the Rotation Algorithm satisfy $\|v_k^{(j)} - V(x_k)\|_H \leq \|v_{k-1} - V(x_k)\|_H$. Since the eigenvalues of H are uniformly bounded below and above, the norms $\|\cdot\|_H, \|\cdot\|$ are equivalent, and hence in particular

$$\|v_k - V(x_k)\| \leq C_7\|v_{k-1} - V(x_k)\| \leq C_7(\|v_{k-1} - v_*\| + \|V(x_k) - v_*\|) = O(h^2 + r)$$

for some constant $C_7 > 0$, since $v_{k-1} \in B_{C_3 r}(v_*)$ and using (44). Combining this with (44) and choosing $h_0^2 \leq r$, we deduce that the Rotation Algorithm terminates with an iterate v_k such that

$$\|v_* - v_k\| \leq \|v_* - V(x_k)\| + \|v_k - V(x_k)\| \leq C_4 r$$

for some constant that depends only on r but is independent of v_{k-1} and remains bounded as $r \to 0$.

At termination the Rotation Algorithm guarantees the estimate

$$\big\|(I - v_k \otimes v_k)H_h(x_k; v_k)\big\| \leq \|\nabla_x \mathcal{E}_h(x_k, v_{k-1})\|.$$

We set $x^t = (1-t)x_h + tx_k$, $v^t = v_h + tv_{k-1}$ and expand

$$\begin{aligned}\big\|\nabla_x \mathcal{E}_h(x_k, v_{k-1})\big\| &= \left\| \int_0^1 \Big(\nabla_x^2 \mathcal{E}_h(x^t, v^t)(x_k - x_h) + \nabla_v \nabla_x \mathcal{E}_h(x^t, v^t)(v_{k-1} - v_h) \Big) \,\mathrm{d}t \right\| \\ &\leq C_2' \big(\|x_k - x_h\| + h^2 \|v_{k-1} - v_h\| \big).\end{aligned}$$

Combined with Lemma 16 this yields the estimate (43).

The statement that $v_k \in B_{C_3 r}(v_*)$ (instead of only $B_{C_4 r}(v_*)$) is an immediate consequence of (43) by ensuring that $C_3 \geq C_2 + C_3 h^2 + C' h^4$, where $\|v_h - v_*\| \leq C' h^2$ for all $h \leq h_0$ from Proposition 2. While there is an interdependence between C_3 and C_2, for r and h_0 sufficiently small, this is clearly achievable. □

We now establish the existence of a minimiser of the auxiliary functional F_k under the conditions ensured by the rotation step of Algorithm 3.

Lemma 18. *Under the conditions of Lemma 17, possibly after choosing a smaller r, h_0, there exists a constant $C_4 > 0$, such that the functional F_k defined in* (12) *has a unique minimiser $y_k \in B_r(x_*)$ satisfying*

$$\|y_k - x_h\| \leq C_4 (r_k^2 + h^2 r_k + h^4 s_{k-1}). \tag{45}$$

Proof. We begin by estimating the residual

$$\nabla F_k(x_h) = \nabla_x \mathcal{E}_h(x_h, v_k) - 2(\nabla_x \mathcal{E}_h(x_k, v_k) \cdot v_k) v_k + 2\lambda_k ((x_k - x_h) \cdot v_k) v_k,$$

where $\lambda_k = H_h(x_k; v_k) \cdot v_k$. We consider each constituent term in this expression in turn; we expand about (x_h, v_h), and use the identities (4), (6) and (22) This gives

$$\begin{aligned}
v_k &= v_h + O(s_k) \\
\nabla_x \mathcal{E}_h(x_h, v_k) &= \nabla_x \nabla_v \mathcal{E}_h(x_h, v_h)(v_k - v_h) + O(s_k^2) \\
\nabla_x \mathcal{E}_h(x_k, v_k) &= \nabla_x^2 \mathcal{E}_h(x_h, v_h)(x_k - x_h) + \nabla_x \nabla_v \mathcal{E}_h(x_h, v_h)(v_k - v_h) + O(r_k^2) + O(s_k^2) \\
&= \nabla_x^2 \mathcal{E}_h(x_h, v_h)(x_k - x_h) + O(h^2 s_k) + O(r_k^2) + O(s_k^2), \\
\nabla_x \mathcal{E}_h(x_k, v_k) \cdot v_k v_k &= (\nabla_x^2 \mathcal{E}_h(x_h, v_h)(x_k - x_h) + \nabla_x \nabla_v \mathcal{E}_h(x_h, v_h)(v_k - v_h)) \cdot v_h v_h \\
&\quad + O(r_k^2) + O(s_k^2) + O(r_k s_k) \\
&= \nabla_x^2 \mathcal{E}_h(x_h, v_h)(x_k - x_h) \cdot v_h v_h + O(h^2 s_k) + O(r_k^2) + O(s_k^2) + O(r_k s_k) \\
H_h(x_k; v_k) &= H_h(x_h; v_h) + O(r_k) + O(s_k) \\
\lambda_k &= \lambda_h + v_k \cdot H_h(x_k; v_k) - v_h \cdot H_h(x_h; v_h) = \lambda_h + O(r_k) + O(s_k) \\
\lambda_k ((x_k - x_h) \cdot v_k) v_k &= (\lambda_h + O(r_k) + O(s_k))((x_k - x_h) \cdot v_k) v_k \\
&= \lambda_h ((x_k - x_h) \cdot v_h) v_h + O(r_k^2) + O(r_k s_k).
\end{aligned}$$

Thus since (7) and our assumption that $v_{k-1} \in B_{C_3 r}(v_*)$ ensure that $s_{k-1} = O(1 + h_0^2)$, while (43) implies that $s_k = O(r_k) + O(h^2 s_{k-1})$, we combine the above to obtain

$$\begin{aligned}\nabla F_k(x_h) = -2\big[(\nabla_x^2 \mathcal{E}_h(x_h, v_h)(x_k - x_h)) \cdot v_h\big] v_h + 2\lambda_h((x_k - x_h) \cdot v_h) v_h \\ + O\big(r_k^2 + h^2 r_k + h^4 s_{k-1}\big),\end{aligned}$$

Next, we note that, by definition of $\mathcal{E}_h$, $\nabla_x^2 \mathcal{E}_h(x_h, v_h) v_h = \nabla^2 E(x_h) v_h + O(h^2)$, and thus from (2) that $\nabla_x^2 \mathcal{E}_h(x_h, v_h) v_h = H_h(x_h; v_h) + O(h^2)$. Hence applying (6),

$$\begin{aligned}\nabla F_k(x_h) &= \big[-2H_h(x_h; v_h) \cdot (x_k - x_h) + 2\lambda_h (x_k - x_h) \cdot v_h\big] v_h + O\big(r_k^2 + h^2 r_k + h^4 s_{k-1}\big) \\ &= O\big(r_k^2 + h^2 r_k + h^4 s_{k-1}\big). \end{aligned} \tag{46}$$

Finally, we observe that $\nabla^2 F_k(x_h)$ is positive definite, since

$$\begin{aligned}\nabla^2 F_k(x_h) &= \nabla_x^2 \mathcal{E}_h(x_h, v_k) - 2\lambda_k v_k \otimes v_k \\ &= \nabla_x^2 \mathcal{E}_h(x_h, v_h) - 2\lambda_h v_h \otimes v_h + O(r_k) \\ &= \nabla^2 E(x_*) - 2\lambda_* v_* \otimes v_* + O(h^2 + r_k), \end{aligned} \tag{47}$$

which immediately implies that, for r, h_0 sufficiently small, $\nabla^2 F_k(x_h)$ is an isomorphism with uniformly bounded inverse.

Thus an application of the inverse function theorem to ∇F_k at y_k using (46) yields the stated result. □

We now turn towards analysing the linesearch for x. Recall the definition of the energy norm $\|u\|_* := \sqrt{u \cdot ((I - 2v_* \otimes v_*) \nabla^2 E(x_*) u)}$, which is equivalent to $\|\cdot\|$. In particular,

$$\mu^{1/2} \|u\| \le \|u\|_* \le \|\nabla^2 E(x_*)\| \|u\| \quad \text{where} \quad \mu := \min(-\lambda_*, \mu_*) > 0. \tag{48}$$

Lemma 19. *There exists $r, h_0, \underline{\alpha} \in (0, \alpha_0]$ and $\gamma_* \in (0, 1)$, such that, if $h \in (0, h_0]$, $x_k \in B_r(x_*), v_k \in B_{C_3 r}(v_*)$ and $\alpha_{k-1} \ge \underline{\alpha}$, then*

$$\alpha_k \ge \underline{\alpha} \qquad \text{and} \qquad \|x_{k+1} - y_k\|_* \le \gamma_* \|x_k - y_k\|_*,$$

where y_k is the minimiser of F_k established in Lemma 18.

Proof. We begin by noting that, for any $r > 0$, the norms $\|\nabla^2 F_k(x)\|$ are uniformly bounded among all choices of $x_k \in B_r(x_*)$, $x \in B_{r+1}(x_*)$. This is straightforward to establish.

Therefore, there exists $\underline{\alpha} > 0$ such that, for $x_k \in B_r(x_*)$ and for any $\alpha \in (0, 2\underline{\alpha}]$, the conditions in Step (6) of Algorithm 3 are met (this includes an Armijo condition for F_k) since ∇F_k is Lipschitz in a neighbourhood of x_k [8, Thm.2.1]. It is no restriction of generality to require $\underline{\alpha} \le \alpha_0$. In particular, $\alpha_k \ge \underline{\alpha}$.

For r, h_0 sufficiently small, we have $y_k \in B_r(x_*)$ as well. Upon choosing r sufficiently small, $u \cdot (\nabla^2 F_k(y) u) \ge \mu/2 \|u\|^2$ for all $u \in X$, $y \in B_r(x_*)$. Thus, we can apply Lemma 14

(with $x_* \equiv y_k$) to deduce that, for r sufficiently small, the step $x_{k+1} = x_k - \alpha_k \nabla F(x_k)$ is a contraction with a constant γ_1 that is independent of x_k, v_k. That is,

$$(x_{k+1} - y_k) \cdot \big[\nabla^2 F_k(y_k)(x_{k+1} - y_k)\big] \leq \gamma_1^2 (x_k - y_k) \cdot \big[\nabla^2 F_k(y_k)(x_k - y_k)\big],$$

Recalling from (45) and (47) that $\nabla^2 F_k(y_k) = (I - 2v_* \otimes v_*)\nabla^2 E(x_*) + O(r + h^2)$ we find that, for r, h_0 sufficiently small,

$$\|x_{k+1} - y_k\|_* \leq \gamma_* \|x_k - y_k\|_*, \tag{49}$$

where $\gamma_* \in [\gamma_1, 1)$, again independent of x_k, v_k, but depending on r, h_0. □

We have now assembled all prerequisites required to complete the proof of Theorem 7.

Inspired by Lemma 19, our aim is to prove that, for r sufficiently small, there exists $\gamma \in (0, 1)$ such that, for all $j \geq 0$,

$$r_j^* + h^2 s_{j-1} \leq \gamma^j (r_0^* + h^2 s_{-1}) =: \gamma^j t_0, \tag{50}$$

where $\gamma := \frac{1}{2}(\gamma_* + 1)$, $r_k^* := \|x_k - x_h\|_*$ and $s_k := \|v_k - v_h\|$.

A consequence of (50) would be that there exists a constant c such that $\|x_j - x_*\| \leq cr =: \hat{r}$. Thus, under the assumptions of the Theorem, let r, h_0 be chosen sufficiently small so that Proposition 2, and Lemmas 16, 17, 18 and 19 apply with r replaced by $\hat{r}$.

We now begin the induction argument adding to (50) the conditions that

$$v_{j-1} \in B_{C_3 r}(v_*) \qquad \text{and} \qquad \alpha_j \geq \underline{\alpha}, \tag{51}$$

where $C_3 \geq 1$ is the constant from Lemma 17 and $\underline{\alpha}$ the constant from Lemma 19. Clearly (50) and (51) hold for $j = 0$. Suppose that they hold for $j = 0, \ldots, k$, where $k \geq 0$.

The choice of r implies that $x_k \in B_r(x_*)$ again, and Lemma 17 implies that $v_k \in B_{C_3 r}(v_*)$. Thus, the first condition in (51) is established for $j = k + 1$.

Applying Lemma 19 we obtain the second condition in (51) for $j = k + 1$, and in addition that

$$\|x_{k+1} - y_k\|_* \leq \gamma_* \|x_k - y_k\|_*,$$

where y_k is the minimiser of F_k established in Lemma 18. Using (49), the fact that $\gamma_* < 1$ and Lemma 18 we therefore deduce that there exists a constant C_5 which depends on C_4 and on the norm-equivalence between $\|\cdot\|$ and $\|\cdot\|_*$, such that

$$\begin{aligned}
\|x_{k+1} - x_h\|_* &\leq \|x_{k+1} - y_k\|_* + \|y_k - x_h\|_* \\
&\leq \gamma_* \|x_k - y_k\|_* + \|y_k - x_h\|_* \\
&\leq \gamma_* \|x_k - x_h\|_* + 2\|y_k - x_h\|_* \\
&\leq (\gamma_* + C_5 h^2 + C_5 r_k)\|x_k - x_h\|_* + C_5 h^4 \|v_{k-1} - v_h\|.
\end{aligned}$$

Adding $h^2 \|v_k - v_h\|$ to both sides of the inequality and applying (43) and (48) we thus obtain

$$\begin{aligned}
r_{k+1}^* + h^2 s_k &\leq (\gamma_* + C_5 h^2 + C_5 r_k) r_k^* + h^2 s_k + C_5 h^4 s_{k-1} \\
&\leq \big(\gamma_* + C_5 h^2 + \mu^{-1/2} C_2 h^2 + C_5 (c+1) r\big) r_k^* + (C_5 + C_2) h^4 s_{k-1}.
\end{aligned}$$

Recalling that $\gamma = \frac{1}{2}(\gamma_* + 1)$, choosing h_0, r sufficiently small, we obtain that

$$r^*_{k+1} + h^2 s_k \leq \gamma(r^*_k + h^2 s_{k-1}).$$

This establishes (50) for $j = k + 1$ and thus completes the induction argument.

In summary, we have proven that (50) and (51) hold for all $j \geq 0$. As a first consequence, we obtain that $r_k := \|x_k - x_h\| \leq \mu^{-1/2}\|\nabla^2 E(x_*)\|\gamma^k(r_0 + h^2 s_{-1})$ using (48), which in particular establishes the first part of (14).

To obtain a convergence rate for v_k we combine (43) and (50), to obtain

$$\|v_k - v_h\| \leq C_6(r^*_k + h^2 s_{k-1}) \leq C_6\gamma^k t_0 \leq C_6\|\nabla^2 E(x_*)\|\gamma^k(r_0 + h^2 s_{-1}),$$

for a constant C_6. Choosing $C = 2\max(C_6, \mu^{-1/2})\|\nabla^2 E(x_*)\|$ completes the proof of Theorem 7.

Remark 20. We can slightly improve the convergence result by also analysing the Rotation step in more detail:

Having established the convergence and in particular boundedness of the x_k, v_k iterates of Algorithm 3 it follows immediately that there exists a lower bound $\underline{\beta}$ such that the Armijo condition in Step (6) of the Rotation Algorithm is satisfied for all k and for all $\beta \in (0, 2\underline{\beta}]$. We can therefore conclude that always $\beta \geq \underline{\beta}$.

With this lower bound, and possibly choosing r smaller than above, Lemma 15 implies that the steepest descent step, Step (8) of the Rotation Algorithm, is a contraction towards $V(x_k)$ with a contraction factor γ_v that is *independent* of k. Thus, for r sufficiently small, the number of gradient and energy evaluations per outer iteration of Algorithm 3 is bounded by a constant m that is independent of k.

This allows us to conclude that there exist constants $C > 0, \rho \in (0, 1)$ such that

$$r_k + s_k \leq C\rho^{\#\mathrm{E}(k)},$$

where $\#\mathrm{E}(k)$ denotes the total number of gradient and energy evaluations carried out in iterations $k = 0$ through k of Algorithm 3 (including the Rotation Algorithm). □

References

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

[2] P.-A. Absil, R. Mahony, and J. Trumpf. An extrinsic look at the Riemannian Hessian. In F. Nielsen and F. Barbaresco, editors, *Geometric Science of Information*, number 8005 in Lecture Notes in Computer Science, pages 361–368, Heidelberg, Berlin, New York, 2013. Springer Verlag.

[3] A. Banerjee, N. Adams, J. Simons, and R. Shepard. Search for stationary points on surfaces. *The Journal of Physical Chemistry*, 89:52–57, 1985.

[4] G.T. Barkema and N. Mousseau. Event-based relaxation of continuous disordered systems. *Physical Review Letters*, 77:4358, 1996.

[5] G.T. Barkema and N. Mousseau. The activation-relation technique: an efficient algorithm for sampling energy landscapes. *Computational Materials Science*, 20(3–4):285–292, 2001.

[6] E. Cances, F. Legoll, M.C. Marinica, K. Minoukadeh, and F. Willaime. Some improvements of the activation-relaxation technique method for finding transition pathways on potential energy surfaces. *The Journal of Chemical Physics*, 130(114711), 2009.

[7] C.J. Cerjan and W.H. Miller. On finding transition states. *Journal of Chemical Physics*, 75:2800, 1981.

[8] N. I. M. Gould and S. Leyffer. An introduction to algorithms for nonlinear optimization. In A. W. Craig J. F. Blowey and T. Shardlow, editors, *Frontiers in Numerical Analysis (Durham 2002)*, pages 109–197, Heidelberg, Berlin, New York, 2003. Springer Verlag.

[9] G. Henkelman and H. Jónsson. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *Journal of Chemical Physics*, 111(5):7010–7022, 1999.

[10] A. Heyden, A. T. Bell, and F. J. Keil. Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method. *Journal of Chemical Physics*, 123(224101), 2005.

[11] H. Jónsson, G. Mills, and K. W. Jacobsen. Nudged elastic band for finding minimum energy paths of transitions. In G. Ciccotti B. J. Berne and D. F. Coker, editors, *Classical and quantum dynamics in condensed phase simulations*, volume 385. World Scientific, 1998.

[12] J. Kästner and P. Sherwood. Superlinearly converging dimer method for transition state search. *The Journal of Chemical Physics*, 128(014106), 2008.

[13] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Journal of research of the National Bureau of Standards B*, 45:225–280, 1950.

[14] D. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming, Series B*, 45(3):503–528, 1989.

[15] E. Machado-Charry, L.K. Beland, D. Caliste, Luigi Genovese, T. Deutsch, N. Mousseau, and P. Pochet. Optimized energy landscape exploration using the ab initio based activation-relaxation technique. *Journal of Chemical Physics*, 135(034102), 2011.

[16] M.C. Marinica, F. Willaime, and N. Mousseau. Energy landscape of small clusters of self-interstitial dumbbells in iron. *Physical Review B*, 83(094119), 2011.

[17] N. Mousseau, L.K. Beland, P. Brommer, J.F. Joly, F. El-Mellouhi, E. Machado-Charry, M.C. Marinica, and P. Pochet. The activation-relaxation technique: Art nouveau and kinetic art. *Journal of Atomic, Molecular and Optical Phsyics*, 2012(952278), 2012.

[18] B. A. Murtagh and R. W. H. Sargent. Computational experience with quadratically convergent minimisation methods. *The Computer Journal*, 13:185–194, 1970.

[19] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999.

[20] R. A. Olsen, G. J. Kroes, G. Henkelman, A. Arnaldsson, and H. Jnsson. Comparison of methods for finding saddle points without knowledge of the final states. *Journal of Chemical Physics*, 121:9776, 2004.

[21] J. Simons, P. Joergensen, H. Taylor, and J. Ozment. Walking on potential energy surfaces. *The Journal of Physical Chemistry*, 87:2745–2753, 1983.

[22] B. P. Uberuaga, F. Montalenti, T. C. Germann, and A. F. Voter. Accelerated molecular dynamics methods. In S. Yip, editor, *Handbook of Materials Modelling, Part A-Methods*, page 629. Springer, 2005.

[23] A. E. Perekatov V. S. Mikhalevich, N. N. Redkovskii. Methods of minimization of functions on a sphere and their applications. *Cybernetics and Systems Analysis*, 23(6):721–730, 1987.

[24] A. F. Voter. Accelerated molecular dynamics of infrequent events. *Physical Review Letters*, 78(3908), 1997.

[25] A. F. Voter. Introduction to the kinetic Monte Carlo method. In K. E. Sickafus, E. A. Kotomin, and B. P. Uberuaga, editors, *Radiation Effects in Solids*, volume 235 of *NATO Science Series*, pages 1–23. Springer Netherlands, 2007.

[26] E. Weinan, W. Ren, and E. Vanden-Eijnden. String method for the study of rare events. *Physical Review B*, 66(052301), 2002.

[27] E. Weinan, W. Ren, and E. Vanden-Eijnden. Simplified and improved string method for computing the minimum energy path in barrier-crossing events. *Journal of Chemical Physics*, 126(164103), 2007.

[28] J. Zhang and Q. Du. Shrinking dimer dynamics and its applications to saddle point search. *SIAM Journal of Numerical Analysis*, 50(4):1899–1921, 2012.