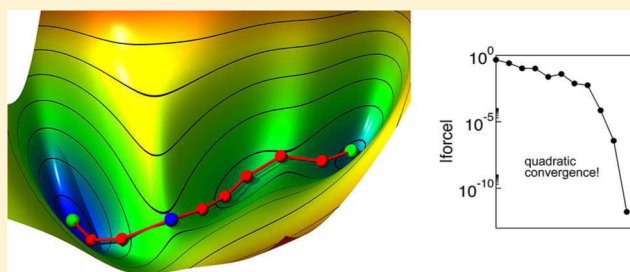


# A Quadratically-Converging Nudged Elastic Band Optimizer

Matthias U. Bohner, Jan Meisner, and Johannes Kästner\*

Computational Biochemistry Group, Institute of Theoretical Chemistry, University of Stuttgart, Pfaffenwaldring 55, 70569 Stuttgart, Germany

**ABSTRACT:** Nudged elastic band (NEB) is a well established chain-of-states method to locate the minimum energy path in configuration space. Unfortunately, existing search algorithms suffer from slow convergence. We provide an analytic derivative of the nudged elastic band force, enabling a full Newton–Raphson optimization. For molecular systems, the components of the step belonging to translations and rotations are removed with an efficient algorithm. Minimization of the NEB force is ensured by reversing components for which the Newton–Raphson step would increase the force. We achieve quadratic convergence of this optimizer when applied to simple test cases where analytic Hessians are available: one analytic two-dimensional potential and a system of Lennard-Jones particles.



## 1. INTRODUCTION

Each configuration of a set of atoms, each geometry, has a certain potential energy. These energies form the potential energy surface in configuration space. Stable states are represented by minima since the system will return to them after small deflections. Reactions are described by a path from one minimum to another. Exploring reaction mechanisms is a key task in computational chemistry and often leads to a deeper understanding of chemical reactions. A mechanism is characterized by its reaction path. Where entropic contributions can be neglected the reaction path with the highest statistical weight is the minimum energy path (MEP). It was shown that the MEP leads over one or more saddle points and is always tangential to the local gradient.<sup>1,2</sup>

Different classes of methods exist to find a MEP. Chain of states methods approximate it as a series of points in configuration space. They include the nudged elastic band<sup>3–6</sup> (NEB) method and variants thereof,<sup>7</sup> string methods,<sup>8–10</sup> and conjugate peak refinement.<sup>11</sup> In this paper, we build on NEB. A good introduction to the NEB method can be found in ref 12.

In NEB, the path is discretized into a number of replicas of the system, called images. The potential energy  $V(R_i)$  is calculated, where  $R_i$  is the position of the  $i$ th image in configuration space. In addition the force  $F_i = -\nabla V(R_i)$  acting on each image is calculated. The aim is to minimize the component of the force perpendicular to the path. The simplest way to do so would be to solve the equation

$$0 \stackrel{!}{=} (1 - \hat{\tau}_i \hat{\tau}_i^T) F_i \quad (1)$$

for each image where  $\hat{\tau}_i$  is the unit vector tangential to the path. The projection operator is  $\hat{\tau} \hat{\tau}^T$ . Several schemes to approximate the tangent vector for discretized paths have been developed.<sup>4,12</sup> Here we stick to the upwind scheme, used in the improved tangent NEB method,<sup>4</sup>

$$\tau_i = \begin{cases} \tau_i^+ = R_{i+1} - R_i & \text{if } V_{i+1} > V_i > V_{i-1} \\ \tau_i^- = R_i - R_{i-1} & \text{if } V_{i+1} < V_i < V_{i-1} \end{cases} \quad (2)$$

where  $V_i = V(R_i)$  is the corresponding potential energy of image  $i$ . If an image has a lower ( $V_{i+1} > V_i < V_{i-1}$ ) or higher ( $V_{i+1} < V_i > V_{i-1}$ ) energy than both its neighbors, a linear combination of both differences is used

$$\tau_i = \begin{cases} \Delta V_i^{\max} \tau_i^+ + \Delta V_i^{\min} \tau_i^- & \text{if } V_{i+1} > V_{i-1} \\ \Delta V_i^{\min} \tau_i^+ + \Delta V_i^{\max} \tau_i^- & \text{if } V_{i+1} < V_{i-1} \end{cases} \quad (3)$$

with

$$\begin{aligned} \Delta V_i^{\max} &= \max(|V_{i+1} - V_i|, |V_{i-1} - V_i|) \\ \Delta V_i^{\min} &= \min(|V_{i+1} - V_i|, |V_{i-1} - V_i|) \end{aligned} \quad (4)$$

In the following, we use the general expression  $\tau_i = a\tau_i^+ + b\tau_i^-$  and chose  $a$  and  $b$  according to the case.

An optimizer based on eq 1 would cause many images to gather near the minima. This problem can be addressed by introducing a penalty, which leads to an equal distribution of the images. In NEB, nonphysical spring forces between the images are used. However, this force, unlike a real spring, should only act tangentially to the path. Altogether in improved-tangent NEB, the force on image  $i$  to be minimized reads<sup>4</sup>

$$F_i^{\text{NEB}} = k(|\tau_i^+| - |\tau_i^-|)\hat{\tau}_i + (1 - \hat{\tau}_i \hat{\tau}_i^T) F_i^t \quad (5)$$

where  $k$  denotes the spring constant. We introduced the superscripts NEB and t to discriminate between the NEB force and the true force  $F^t = -\nabla V$ . The spring, or penalty, term  $k(|\tau_i^+|$

Received: April 19, 2013

Published: July 3, 2013



$-|\tau_i^-|\hat{\tau}_i$  reduces the difference between the distances to both neighbor images but not the distances themselves. This is necessary to avoid artificially shortening the NEB-path (corner cutting). The penalty term should act on physically meaningful differences between the images. Thus, molecular geometries have to be superimposed to remove any translation or rotation between the images.<sup>12</sup>

The key task of NEB optimizations is to minimize the force given by eq 5. Previously, this was done by comparatively slowly converging optimization schemes such as steepest descent or conjugate gradient methods. Better converging optimizers for NEB optimizations are the adopted basis Newton–Raphson scheme,<sup>6</sup> the related limited-memory quasi-Newton (L-BFGS)<sup>13,14</sup> and slow-response quenched velocity Verlet (SQVV) minimizers.<sup>5</sup> Approximations of the derivative of the NEB force can either be obtained for the whole path, or image by image.<sup>13</sup> The latter approach neglects all couplings between the images. We introduce a quadratically converging NEB optimizer based on the Newton–Raphson method by calculating the full analytic derivative of the NEB force in the following.

This paper is organized as follows. In section 2.1 the derivative of the NEB force with respect to the coordinates of the images is derived. This is required to form a Newton–Raphson step. The global translation and rotation of the NEB path are dealt with in section 2.2. The convergence properties of the derived algorithm are then demonstrated in section 3 for an analytic two-dimensional example, as well as at a Lennard–Jones cluster.

## 2. METHOD

**2.1. Derivative of the NEB Force.** The root of a multidimensional function  $F(R)$  can be iteratively approached using the Newton–Raphson method

$$R_{n+1} = R_n - \mathbf{J}^{-1}F(R) \quad (6)$$

where  $n$  is the number of the iteration and  $\mathbf{J}$  is the Jacobian matrix of the function. In our case, we are looking for the root of the NEB force, eq 5. In many optimization problems, a scalar potential function can be found, the minimum of which is sought in the optimization. Then, the Hessian, the symmetric matrix of second derivatives of the potential function, can be used in Newton–Raphson optimizations. However, such a potential is not available for the NEB force, which is why we have to use the Jacobian matrix, which is, in general, not symmetric.

We define an  $(N \cdot M)$ -dimensional super vector, where  $N$  is the number of images and  $M$  the number of coordinates in each image. For a molecule with 5 atoms  $M$  will in general be  $M = 15$ . This super vector reads  $F^{\text{NEB}} = ((F_1^{\text{NEB}})^T, (F_2^{\text{NEB}})^T, (F_3^{\text{NEB}})^T, \dots, (F_N^{\text{NEB}})^T)^T$ . We divide the Jacobian into submatrices

$$\mathbf{J}^{\text{NEB}} = \begin{pmatrix} \mathbf{J}_{1,1}^{\text{NEB}} & \mathbf{J}_{1,2}^{\text{NEB}} & \dots & \mathbf{J}_{1,N}^{\text{NEB}} \\ \mathbf{J}_{2,1}^{\text{NEB}} & \mathbf{J}_{2,2}^{\text{NEB}} & \dots & \mathbf{J}_{2,N}^{\text{NEB}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{J}_{N,1}^{\text{NEB}} & \mathbf{J}_{N,2}^{\text{NEB}} & \dots & \mathbf{J}_{N,N}^{\text{NEB}} \end{pmatrix} \quad (7)$$

with the vector gradient

$$\mathbf{J}_{i,j}^{\text{NEB}} = \mathbf{grad}_{F_i} F_i^{\text{NEB}} = (\nabla_{R_i} \otimes F_i^{\text{NEB}})^T \quad (8)$$

where the derivative of the NEB force of image  $i$  with respect to the coordinates of image  $j$  is calculated and  $x \otimes y = xy^T$  denotes the dyadic product.  $\mathbf{J}_{ij}^{\text{NEB}}$  is an  $M \times M$  matrix.

To derive the Jacobian of the NEB force, eq 5, the following general expressions will be used.

$$\mathbf{grad} A = (\nabla \otimes A)^T \quad (9)$$

$$\mathbf{grad}(gA) = A \otimes (\nabla g) + [g(\nabla \otimes A)]^T \quad (10)$$

$$\mathbf{grad}(A \cdot B) = (\mathbf{grad}(A))^T B + (\mathbf{grad}(B))^T A \quad (11)$$

Here  $A$  and  $B$  are vector-valued and  $g$  is a scalar-valued function.

A further ingredient is the gradient of the length of the tangent vector

$$\mathbf{grad}_j |\tau_i^{+/-}| = \frac{1}{2} \frac{1}{|\tau_i^{+/-}|} \mathbf{grad}_j (\tau_i^{+/-})^2 \quad (12)$$

which leads with eq 11 to

$$\mathbf{grad}_j |\tau_i^+| = \frac{1}{|\tau_i^+|} (\delta_{i+1}^j - \delta_i^j) \tau_i^+ = \hat{\tau}_i^+ (\delta_{i+1}^j - \delta_i^j) \quad (13)$$

and, equivalently

$$\mathbf{grad}_j |\tau_i^-| = \hat{\tau}_i^- (\delta_i^j - \delta_{i-1}^j) \quad (14)$$

with  $\delta$  being the Kronecker delta. Further indispensable relations are

$$-\mathbf{grad}_j F_i^t = \mathbf{H}_i^t \delta_i^j = \mathbf{H}_i^{tT} \delta_i^j \quad (15)$$

$$\hat{\tau}_i \otimes ((\mathbf{grad}_j \hat{\tau}_i) F_i^t) = (\hat{\tau}_i \otimes F_i^t) \mathbf{grad}_j \hat{\tau}_i \quad (16)$$

$\mathbf{H}_i^t$  denotes the Hessian of the true potential  $V(R_i)$ . A straightforward calculation leads then to

$$\begin{aligned} \mathbf{J}_{i,j}^{\text{NEB}} = & -k \hat{\tau}_i \otimes \{\hat{\tau}_+ (\delta_{i+1}^j - \delta_i^j) - \hat{\tau}_- (\delta_i^j - \delta_{i-1}^j)\} + \mathbf{H}_i^t \delta_i^j \\ & - \hat{\tau}_i \mathbf{H}_i^t \hat{\tau}_i^j + \{(F_i^t \cdot \hat{\tau}_i) \mathbb{I} + (\hat{\tau}_i \otimes F_i^t) - k(|\tau_+| - |\tau_-|) \mathbb{I}\} \mathbf{grad}_j \hat{\tau}_i \end{aligned} \quad (17)$$

The vector gradient of the unit tangent vector reads

$$\mathbf{grad}_j \hat{\tau}_i = |\tau_i|^{-1} (\mathbb{I} - \hat{\tau}_i \otimes \hat{\tau}_i) \mathbf{grad}_j \tau_i \quad (18)$$

Plugging it into eq 17 gives

$$\begin{aligned} \mathbf{J}_{i,j}^{\text{NEB}} = & -k \hat{\tau}_i \otimes \{\hat{\tau}_+ (\delta_{i+1}^j - \delta_i^j) - \hat{\tau}_- (\delta_i^j - \delta_{i-1}^j)\} + \mathbf{H}_i^t \delta_i^j \\ & - \hat{\tau}_i \mathbf{H}_i^t \hat{\tau}_i^j + \mathbf{A} \cdot \mathbf{grad}_j \tau_i \end{aligned} \quad (19)$$

with

$$\mathbf{A} = \{[F_i^t \cdot \hat{\tau}_i - k(|\tau_+| - |\tau_-|)] \mathbb{I} + (\hat{\tau}_i \otimes F_i^t)\} \cdot \frac{\mathbb{I} - \hat{\tau}_i \otimes \hat{\tau}_i}{|\tau_i|} \quad (20)$$

Now we can see, that the Jacobian of the NEB force of eq 5 is a band matrix since every term contains one of the Kronecker deltas  $\{\delta_{i-1}^j, \delta_i^j, \delta_{i+1}^j\}$ . However, keep in mind that the indices  $i$  and  $j$  do not determine a single element but an  $M \times M$  submatrix.

While  $\hat{\tau}_i$  depends on the following case distinctions, eq 20 provides  $\mathbf{A}$  for all cases.

**Case:**  $V_{i+1} > V_i > V_{i-1}$ .  $\tau_i = \tau_i^+ = R_{i+1} - R_i$ . The vector gradient  $\mathbf{grad}_j \tau_i^+$  reads

$$\mathbf{grad}_j \tau_i = \mathbf{grad}_j \tau_i^+ = (\delta_{i+1}^j - \delta_i^j) \mathbb{I} \quad (21)$$

We get the diagonal and subdiagonal elements of the NEB–Jacobian

$$\mathbf{J}_{i,i-1}^{\text{NEB}} = -k\hat{\tau}_i \otimes \hat{\tau}_i^- \quad (22)$$

$$\mathbf{J}_{i,i}^{\text{NEB}} = k\hat{\tau}_i \otimes \{\hat{\tau}_i^+ + \hat{\tau}_i^-\} + \mathbf{H}_i - \hat{\tau}_i \mathbf{H}_i \hat{\tau}_i - \mathbf{A} \quad (23)$$

$$\mathbf{J}_{i,i+1}^{\text{NEB}} = -k\hat{\tau}_i \otimes \hat{\tau}_i^+ + \mathbf{A} \quad (24)$$

All other elements in the  $i^{\text{th}}$  row vanish.

*Case:*  $V_{i+1} < V_i < V_{i-1}$ .  $\tau_i = \tau^- = R_i - R_{i-1}$ . The vector gradient  $\mathbf{grad}_j \tau_i^-$  reads

$$\mathbf{grad}_j \tau_i = \mathbf{grad}_j \tau_i^- = (\delta_i^j - \delta_{i-1}^j) \mathbb{I} \quad (25)$$

We get the diagonal and subdiagonal elements of the NEB–Jacobian

$$\mathbf{J}_{i,i-1}^{\text{NEB}} = -k\hat{\tau}_i \otimes \hat{\tau}_i^- - \mathbf{A} \quad (26)$$

$$\mathbf{J}_{i,i}^{\text{NEB}} = k\hat{\tau}_i \otimes \{\hat{\tau}_i^+ + \hat{\tau}_i^-\} + \mathbf{H}_i - \hat{\tau}_i \mathbf{H}_i \hat{\tau}_i + \mathbf{A} \quad (27)$$

$$\mathbf{J}_{i,i+1}^{\text{NEB}} = -k\hat{\tau}_i \otimes \hat{\tau}_i^+ \quad (28)$$

All other elements in the  $i^{\text{th}}$  row vanish.

*Other Cases.* These cover the situation of image  $i$  being a maximum or a minimum along the path. Then,  $\tau_i$  is a linear combination

$$\tau_i = a\tau^+ + b\tau^- = aR_{i+1} + (b-a)R_i - bR_{i-1} \quad (29)$$

The vector gradient  $\mathbf{grad}_j \tau_i$  reads

$$\begin{aligned} \mathbf{grad}_j \tau_i &= (a\delta_{i+1}^j + (b-a)\delta_i^j - b\delta_{i-1}^j) \mathbb{I} + \tau_i^+ \otimes \mathbf{grad}_j(a) \\ &\quad + \tau_i^- \otimes \mathbf{grad}_j(b) \end{aligned} \quad (30)$$

*Case:*  $a = |V_{i+1} - V_i|$ ,  $b = |V_{i-1} - V_i|$ .

$$\begin{aligned} \mathbf{grad}_j(a) &= \frac{1}{2} \frac{1}{|V_{i+1} - V_i|} 2(V_{i+1} - V_i) \cdot \\ &\quad (\delta_{i+1}^j \nabla V_{i+1} - \delta_i^j \nabla V_i) \\ &= \text{sign}(V_{i+1} - V_i) (\delta_{i+1}^j \nabla V_{i+1} - \delta_i^j \nabla V_i) \end{aligned} \quad (31)$$

$$\mathbf{grad}(b) = \text{sign}(V_{i-1} - V_i) (\delta_{i-1}^j \nabla V_{i-1} - \delta_i^j \nabla V_i) \quad (32)$$

Since the images  $i$  considered here are always minima or maxima  $\text{sign}_i = \text{sign}(V_{i+1} - V_i) = \text{sign}(V_{i-1} - V_i)$

$$\begin{aligned} \mathbf{grad}_j \tau_i &= (a\delta_{i+1}^j + (b-a)\delta_i^j - b\delta_{i-1}^j) \mathbb{I} \\ &\quad + \text{sign}_i \cdot (\delta_{i+1}^j \tau^+ \otimes \nabla V_{i+1} - \delta_i^j (\tau^+ + \tau^-) \\ &\quad \otimes \nabla V_i + \delta_{i-1}^j \tau^- \otimes \nabla V_{i-1}) \end{aligned} \quad (33)$$

The resulting elements of the Jacobian are

$$\mathbf{J}_{i,i-1}^{\text{NEB}} = -k\hat{\tau}_i \otimes \hat{\tau}_i^- - \mathbf{A}\{b\mathbb{I} - \text{sign}_i \cdot \tau^- \otimes \nabla V_{i-1}\} \quad (34)$$

$$\begin{aligned} \mathbf{J}_{i,i}^{\text{NEB}} &= k\hat{\tau}_i \otimes \{\hat{\tau}_i^+ + \hat{\tau}_i^-\} + \mathbf{H}_i - \hat{\tau}_i \mathbf{H}_i \hat{\tau}_i + \mathbf{A}\{(b-a)\mathbb{I} \\ &\quad - \text{sign}_i \cdot (\tau^+ + \tau^-) \otimes \nabla V_i\} \end{aligned} \quad (35)$$

$$\mathbf{J}_{i,i+1}^{\text{NEB}} = -k\hat{\tau}_i \otimes \hat{\tau}_i^+ + \mathbf{A}\{a\mathbb{I} + \text{sign}_i \cdot \tau^+ \otimes \nabla V_{i+1}\} \quad (36)$$

*Case:*  $a = |V_{i-1} - V_i|$ ,  $b = |V_{i+1} - V_i|$ . A similar calculation leads to

$$\mathbf{J}_{i,i-1}^{\text{NEB}} = -k\hat{\tau}_i \otimes \hat{\tau}_i^- - \mathbf{A}\{b\mathbb{I} - \text{sign}_i \cdot \tau^+ \otimes \nabla V_{i-1}\} \quad (37)$$

$$\begin{aligned} \mathbf{J}_{i,i}^{\text{NEB}} &= k\hat{\tau}_i \otimes \{\hat{\tau}_i^+ + \hat{\tau}_i^-\} + \mathbf{H}_i - \hat{\tau}_i \mathbf{H}_i \hat{\tau}_i \\ &\quad + \mathbf{A}\{(b-a)\mathbb{I} - \text{sign}_i \cdot (\tau^+ + \tau^-) \otimes \nabla V_i\} \end{aligned} \quad (38)$$

$$\mathbf{J}_{i,i+1}^{\text{NEB}} = -k\hat{\tau}_i \otimes \hat{\tau}_i^+ + \mathbf{A}\{a\mathbb{I} + \text{sign}_i \cdot \tau^- \otimes \nabla V_{i+1}\} \quad (39)$$

These expressions cover all cases, so that the Jacobian, eq 7, can be set up and the Newton–Raphson method using eq 6 can be implemented.

The first challenge to face is that the radius of convergence of the Newton–Raphson algorithm is, from experience, often rather small. By contrast, Newton–Raphson steps become large for a flat energy surface. Therefore, we adopt a truncated Newton–Raphson approach. We predefine a maximal step length  $d_{\text{max}}$ . If the step length  $d = |\mathbf{J}^{-1}\mathbf{F}|$  turns out larger than  $d_{\text{max}}$  the step is scaled back and we get the iteration formula

$$R_{n+1} = R_n - \mathbf{J}^{-1} \mathbf{F} \frac{d_{\text{max}}}{d} \quad (40)$$

**2.2. Global Translation and Rotation.** For free (i.e., nonfrozen) end points, the NEB force is invariant with respect to translation and rotation of the whole path if the potential energy is invariant under these transformations (which is the case for molecules). Thus, the Jacobian in Cartesian coordinates becomes singular. A step like eq 6 cannot be formed since  $\mathbf{J}$  cannot be inverted. A singular value decomposition (SVD) of the form  $\mathbf{U}^T \mathbf{J} \mathbf{V} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_{\text{NM}})$  of  $\mathbf{J}$  will result in one or more singular values  $\sigma_i$  being zero.

Even if the end points are frozen, the first derivative of the NEB force is zero along the direction of a global translation of the nonfrozen path as long as the frozen end points have a lower energy than their neighboring points. In that case, the upwind scheme, eq 2, causes the tangent vectors at the image next to the frozen end points to point toward the movable neighbor images. Specifically, if image 1 is frozen and the energy increases for the images 1, 2, and 3 in that order, then  $\tau_2 = \tau_2^+ = R_3 - R_2$ . The only term in the expression of the NEB force of image 2 that changes under a global translation of the path is  $|\tau_2^-|$ , see eq 5. However,  $\mathbf{grad}_2 |\tau_2^-| = \hat{\tau}_2^-$ , see eq 14. Since the molecules of the start and end images (and, thus, all intermediate images) are superimposed to remove a translation between them,  $\hat{\tau}_2^-$  is orthogonal to a translation of  $R_2$  and the global NEB force vanishes. The same arguments hold for the other end of the NEB path. It should be noted that the NEB force is only invariant to a global translation of first order. The second derivative is nonzero and, in fact, the NEB force changes for nonzero displacements of the path. For a three-dimensional system, three singular values are zero (two for a two-dimensional system). They correspond to translations. Those corresponding to rotations only approach zero for a converged NEB path with all forces approaching zero. For a system with one intermediate local minimum, the same arguments can be used to prove that translations of the two individual parts of the paths against each other also result in a zero first derivative of the NEB force. Thus, in two dimensions, four singular values are zero for any general path, 6 for the converged path. In the system studied later, three barriers with

two minima between them in two dimensions, six singular values are zero for a general path, 13 for the converged one. Overall, a scheme is needed to deal with singular Jacobian matrices.

The inversion of the Jacobian matrix is replaced by a pseudo inversion of the form

$$\mathbf{J}^{-1} \rightarrow \mathbf{J}^{+} = \mathbf{V} \text{diag} \left( \frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \frac{1}{\sigma_3}, \dots, \frac{1}{\sigma_K} \right) \mathbf{U}^T \quad (41)$$

where  $\mathbf{V}$  and  $\mathbf{U}$  are the matrices of the right-singular and left-singular vectors and  $\sigma_1 \dots \sigma_K$  with  $K \leq \text{NM}$  are all nonzero singular values. Since numerical noise affects the singular values, we ignore those with an absolute value smaller than a threshold.

The aim is to minimize the NEB force. The Newton–Raphson algorithm, however, converges to any stationary point, not necessarily to a minimum. We check the step direction for each pair of singular vectors individually. For the step to point to a smaller NEB force, the scalar product of the right-side singular vector and the left-side singular vector to the same singular value should be positive. If it is negative, we reverse the step in the corresponding direction, that is, we multiply the corresponding singular values by  $-1$ . Practical application shows that this procedure ensures the convergence of the algorithm to the correct MEP.

**2.3. Order of Convergence.** A series  $\{u_k\}$ ,  $k = 1, 2, \dots$  converges with the order  $p$  to its limit  $u_\infty$  if

$$\lim_{k \rightarrow \infty} \frac{|u_{k+1} - u_\infty|}{|u_k - u_\infty|^p} = c \text{ with } c > 0 \quad (42)$$

While this is a formal definition, it is less useful for numerical purposes. We therefore define a series  $\varepsilon_k > 0$  which converges to zero and fulfills condition eq 42. Doing so we call a numerical series to converge with the order  $p$  if

$$|u_k - u_\infty| \leq \varepsilon_k \quad (43)$$

An example for a quadratically ( $p = 2$ ) convergent series is

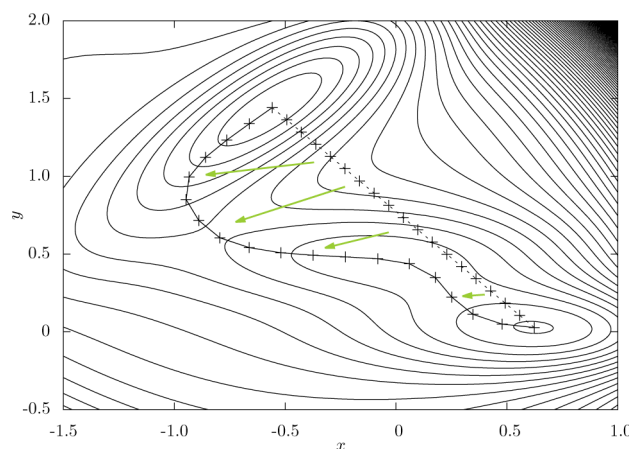
$$\varepsilon_k = 10^{-2^k} \quad (44)$$

Colloquially, it is often said that a series converges quadratically if the number of correct decimal figures doubles with each iteration. However, this is a sufficient, but not a necessary condition.

The modified Newton–Raphson scheme for NEB optimizations presented here was implemented in a local branch of DL-FIND,<sup>15</sup> which is one of the optimizer options in ChemShell.<sup>16</sup> In the examples in the following section, the method presented here is compared to the existing optimizer in DL-FIND, which was used in several applications.<sup>14,17–21</sup> There, a quasi-Newton scheme is applied with the inverse Jacobian obtained from an L-BFGS update<sup>22,23</sup> of a scaled unitary matrix using the NEB force of the whole path. Details of that implementation are given elsewhere,<sup>14</sup> it is similar to an approach termed GL-BFGS previously.<sup>13</sup>

### 3. EXAMPLES

**3.1. Müller–Brown Potential.** We applied the method to the two-dimensional Müller–Brown (MB) potential.<sup>24</sup> A straight line between the two main minima was used as initial guess, see Figure 1. We used 19 images to approximate the path. The first and last image were fixed to the minima. This system is not invariant with respect to translation and rotation,



**Figure 1.** Contour plot of the MB potential. The dashed line shows the initial guess of the NEB path. The solid line shows the converged MEP. Images used to approximate the path are marked by crosses.

since the potential energy changes in each of the two dimensions. Therefore, a direct inversion of the Jacobian was used here rather than the pseudo inversion via the SVD.

The convergence properties of our method are given in Table 1 and compared to the L-BFGS optimizer in Figure 2. Beginning from iteration 7 we achieve quadratic convergence. To demonstrate the convergence rate numerically we compare the remaining  $|F^{\text{NEB}}|$  to a series, which is known to converge quadratically. Sets of parameters other than those used here ( $k = 2.93$ ,  $d_{\text{max}} = 0.15$ ) also lead to quadratic convergence, which may, however, set in only after more cycles.

Table 1 also attempts to quantify how asymmetric the Jacobian actually is. For that purpose, the Euclidean norm of the Jacobian  $\|\mathbf{J}^{\text{NEB}}\|$  is compared to  $\|\mathbf{J}^{\text{NEB}} - (\mathbf{J}^{\text{NEB}})^T\|$ . For a symmetric Jacobian, the latter should be zero. It is clear that the Jacobian significantly deviates from a symmetric matrix. This is one reason why the L-BFGS approach, which results in a symmetric approximative Jacobian matrix, converges slowly compared to the approach shown in this work.

In cases where the step was scaled back according to eq 40, the scaling factor  $d_{\text{max}}/d$  is given in Table 1. Scaling was only necessary before the region of quadratic convergence was entered.

The results clearly show that the optimization algorithm presented in this work efficiently finds the region of convergence of the Newton–Raphson method where it adopts quadratic convergence. This two-dimensional example does not include translational or rotational invariance, however. The resulting Jacobian can be inverted. A more stringent test for molecular systems is presented in the following example.

**3.2. Rearrangement of a Lennard-Jones Cluster.** As a second example we studied the rearrangement of a planar Lennard-Jones cluster of seven particles. The pair interaction is defined by

$$V(r) = 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \quad (45)$$

where  $r$  is the distance between two particles. The parameters  $\varepsilon$  and  $\sigma$  denote the depth and position of the minimum, respectively.

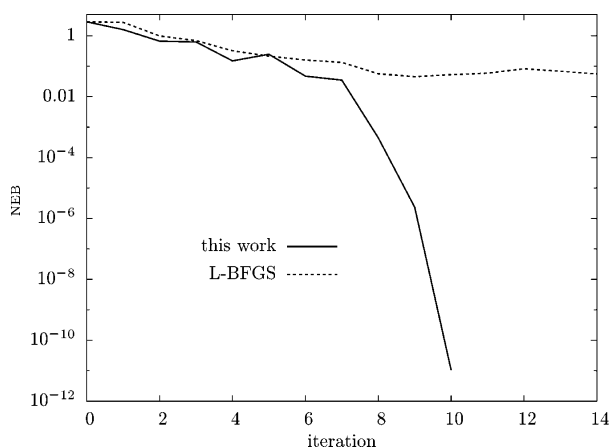
The example was taken from ref 25. The minima along the path are illustrated in Figure 3. A linear interpolation between



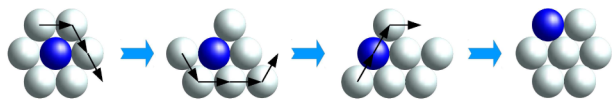
Table 1. Convergence of an NEB Optimization in the MB Potential<sup>a</sup>

iteration	$ F^{\text{NEB}} $	$\epsilon_0/10^{2^k}$	$d_{\text{max}}/d$	$\ J^{\text{NEB}}\ $	$\ J^{\text{NEB}} - (J^{\text{NEB}})^T\ $
0	2.85952		0.39	55.7	48.1
1	1.59505		0.81	53.2	30.8
2	0.66294		0.22	49.9	23.6
3	0.630337		0.93	49.9	21.9
4	0.149348		0.85	51.7	23.2
5	0.247031			53.7	24.4
6	0.047369			54.5	23.6
7	0.0349786	0.1		55.3	23.9
8	0.000450947	0.01		55.3	23.7
9	0.00000235723	0.0001		55.3	23.7
10	0.000000000105215	0.00000001			

<sup>a</sup>The third column shows a suitable series to demonstrate quadratic convergence. Doubling correct digits are bold.



**Figure 2.** Convergence of an NEB optimization in the MB potential with 19 images using the algorithm presented in this work (solid line) compared with the L-BFGS optimizer (dashed line).

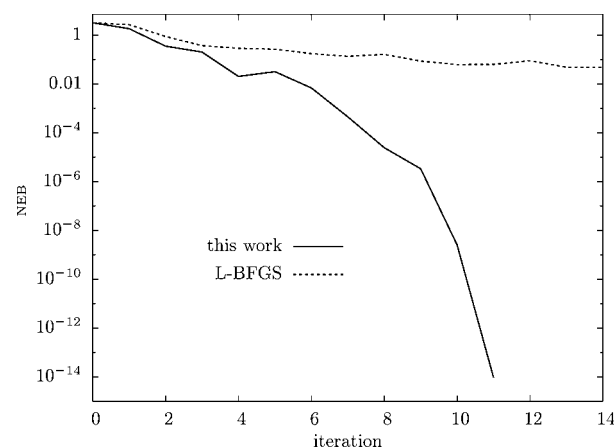


**Figure 3.** Rearrangement of a planar Lennard-Jones cluster with seven atoms. The configurations show the minima which are passed along the path.

these minima was used as initial path. Movement of all atoms was restricted to a two-dimensional plane. This results in 14 degrees of freedom per image including one for rotation and two for translation. Again, 19 images were used to resolve the path unless noted otherwise and the first and last images were frozen. The spring constant was  $k = 15$  and the step limitation  $d_{\text{max}} = 0.04$ . All quantities for this example are given in Lennard-Jones units ( $\epsilon$  and  $\sigma$ ). The images of the initial path were superimposed to minimize rotational and translational displacement. This had been shown previously to be important in NEB optimizations.<sup>12</sup>

For this system, the convergence properties of the optimization method presented here is compared to the L-BFGS algorithm in Figure 4. The resulting numbers are given in Table 2. From iteration 7 onward we obtain quadratic convergence as shown by comparison to a quadratically converging series.

The method converges for a wide range of parameters ( $k$ ,  $d_{\text{max}}$ ), although the number of steps needed to arrive in the basin for quadratic convergence differs. Using more images to



**Figure 4.** Convergence of an NEB optimization for the Lennard-Jones cluster with 19 images using the algorithm presented in this work (solid line) compared with the L-BFGS optimizer (dashed line).

discretize the path shows a difference from the previous example. Here, translations and rotations leave the potential energy unchanged. These additional degrees of freedom transform the intrinsic minima from points to three-dimensional hyperplanes in configurational space. The system can minimize the NEB force by gathering images at a minimum, displaced by translation or rotation. While there should not be any force which displaces images along the translational and rotational modes, numerical errors in the force or in the initial coordinates can build up because of incomplete projection. To improve the stability, we also removed remaining translational components from each Newton–Raphson step before performing it. In Figure 5, we show the energy profile along the NEB path after an NEB optimization with 49 images. Obviously, fewer images are gathered at the minima when the translational component is projected out. In Figure 6, we compare the norm of the total NEB force of all images with altered and unaltered steps. Quadratic convergence is reached in both cases, but it reaches much earlier if we remove translations. Further improvements could be possible by removing rotational components from the steps as well.

Here, we presented calculations for a two-dimensional cluster because it is easier to visualize and to interpret than a real three-dimensional system. However, no further challenges are expected for three-dimensional chemical applications, which will be investigated in further studies.

It should be noted that the invariance of the potential with translation and rotation does not cause problems in the L-

Table 2. Convergence of an NEB Optimization for the Lennard-Jones Cluster Rearrangement with 19 Images

iteration	$ F^{\text{NEB}} $	$\epsilon_0/10^{2^t}$	$d_{\text{max}}/d$	$\ J^{\text{NEB}}\ $	$\ J^{\text{NEB}} - (J^{\text{NEB}})^T\ $
0	3.21598		0.49	318.5	57.3
1	1.8302			273.3	41.3
2	0.352492		0.45	226.8	28.5
3	0.202129			221.5	27.0
4	0.0202993			218.6	26.8
5	0.0318344			217.6	26.5
6	0.00687046			218.4	26.7
7	0.000448025			218.5	26.7
8	0.0000244796	0.0001		218.5	26.7
9	0.00000335986	0.00001		218.5	26.7
10	0.0000000250889	0.0000001		218.5	26.7
11	0.000000000000923465	0.0000000001			

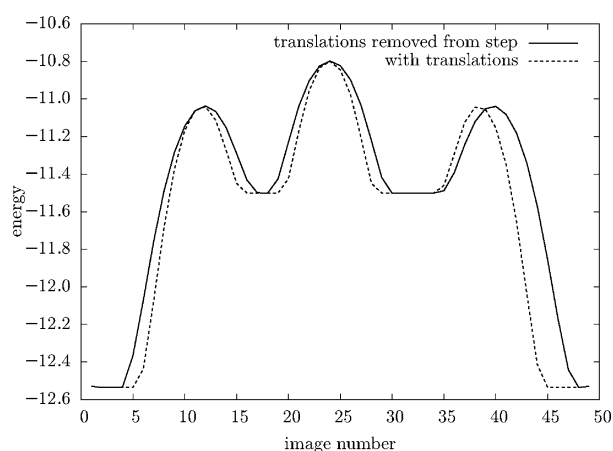


Figure 5. Energy profiles of the NEB paths of the planar Lennard-Jones system with translations removed from each optimization step (solid line) and unaltered steps (dashed).

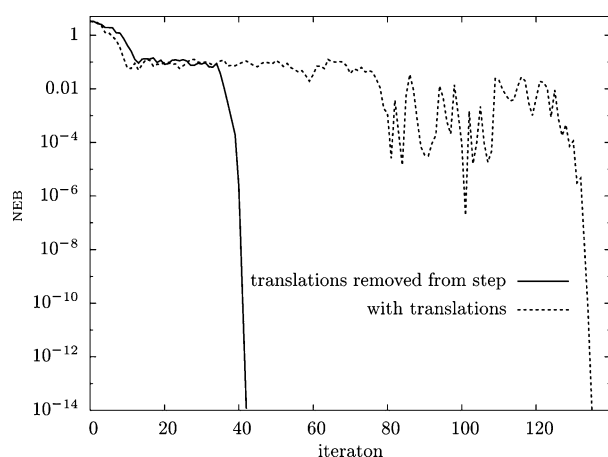


Figure 6. Comparison of the convergence with translations removed from each step (solid line) and unaltered steps (dashed) for the NEB optimization for the Lennard-Jones cluster with 49 images.

BFGS approach. BFGS results in a positive definite Hessian even if the exact one would be positive semidefinite. Translations and rotations are simply ignored in the L-BFGS approach without altering the convergence properties.

#### 4. DISCUSSION

We improved convergence properties of nudged elastic band optimizations by using the analytic derivative of the NEB force

for a Newton–Raphson optimization scheme. Quadratic convergence was demonstrated for two test cases where analytic Hessians of the potential energy are available. The method is not restricted to analytic Hessians. However, we expect a somewhat deteriorated convergence rate for approximate Hessians. In practice in chemical systems, the calculation of analytic Hessians on each point along the path may be considered too time-consuming. Update schemes like the Powell<sup>26</sup> or Bofill<sup>27</sup> updates can be used; however, they are already frequently used in geometry minimizations. Note, that the BFGS update should not be used for Hessians of the individual images, since it results in positive definite matrices by construction, while the Hessians of some of the images have one negative eigenvalue at convergence. These allow one to start from a simple approximation of the Hessian at each image and improve it when more and more gradients at different points in configuration space are calculated. However, the use of updated, approximative Hessians is outside the scope of this work. Here, we provide the exact, analytic derivative of the spring forces and the projections, the nudging.

The computational effort for the SVD scales cubically with the system size ( $N \cdot M$ ). However, for typical quantum chemical problems it is negligible compared to the effort required to calculate the gradients. For larger systems and faster energy expressions, linear scaling optimization algorithms like L-BFGS might be preferable.

In systems invariant to translations and rotations, NEB images tend to accumulate at local minima. These images are only displaced by translations or rotations. This leads to a worse sampling of the areas around the transition states, which we are most interested in. In addition it slows down the convergence. Note that this is a known problem of nudged elastic band calculations in general<sup>12</sup> and not caused by our optimizer. We addressed this issue by removing the translational part from each optimization step. An alternative is an energy penalty for translation and rotation. The problem, obviously, only occurs in molecular systems. In systems with frozen atoms, typically in surface and bulk models, there is no translational or rotational symmetry, which facilitates such NEB optimizations.<sup>13,28</sup>

#### 5. CONCLUSION

We developed a nudged elastic band optimizer based on the analytic derivative of the NEB force with respect to the coordinates of the images along the NEB path. This derivative is used for a Newton–Raphson optimization, which we demonstrated to converge quadratically if analytic Hessians of

the potential energy are available. For translationally or rotationally invariant problems, a pseudoinverse of the Jacobian of the NEB force has to be used. Still, numerical noise may lead to displacements of the NEB images along the translational or rotational components. Therefore, these should be projected out of the optimization step to ensure proper sampling of the minimum-energy path as well as proper convergence.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: kaestner@theochem.uni-stuttgart.de.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Hannes Jónsson for stimulating discussions. This work was financially supported by the German Research Foundation (DFG) within the collaborative research center SFB 716/C.6.

## REFERENCES

- (1) Fukui, K. *Acc. Chem. Res.* **1981**, *14*, 363–368.
- (2) Fukui, K. *Int. J. Quantum Chem.* **1981**, *20*, 633–642.
- (3) Jónsson, H.; Mills, G.; Jacobsen, K. W. Nudged elastic band method for finding minimum energy paths of transitions. *Classical and Quantum Dynamics in Condensed Phase Simulations*; World Scientific, 1998; p 385.
- (4) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9978.
- (5) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9901.
- (6) Chu, J.-W.; Trout, B. L.; Brooks, B. R. *J. Chem. Phys.* **2003**, *119*, 12708.
- (7) Plessow, P. *J. Chem. Theory Comput.* **2013**, *9*, 1305–1310.
- (8) E, W.; Ren, W.; Vanden-Eijnden, E. *Phys. Rev. B* **2002**, *66*, 052301.
- (9) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. *J. Chem. Phys.* **2004**, *120*, 7877.
- (10) Burger, S. K.; Yang, W. *J. Chem. Phys.* **2006**, *124*, 054109.
- (11) Fischer, S.; Karplus, M. *Chem. Phys. Lett.* **1992**, *194*, 252.
- (12) Gonzalez-Garcia, N.; Pu, J.; Gonzalez-Lafont, A.; Lluch, J. M.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 895.
- (13) Sheppard, D.; Terrell, R.; Henkelman, G. *J. Chem. Phys.* **2008**, *128*, 134106.
- (14) Goumans, T. P. M.; Catlow, C. R. A.; Brown, W. A.; Kästner, J.; Sherwood, P. *Phys. Chem. Chem. Phys.* **2009**, *11*, 5431.
- (15) Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. *J. Phys. Chem. A* **2009**, *113*, 11856.
- (16) Sherwood, P.; et al. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 1.
- (17) Metz, S.; Thiel, W. *J. Am. Chem. Soc.* **2009**, *131*, 14885–14902.
- (18) Kästner, J.; Sherwood, P. *Mol. Phys.* **2010**, *108*, 293.
- (19) Rommel, J. B.; Goumans, T. P. M.; Kästner, J. *J. Chem. Theory Comput.* **2011**, *7*, 690.
- (20) Rommel, J.; Liu, Y.; Werner, H.-J.; Kästner, J. *J. Phys. Chem. B* **2012**, *116*, 13682.
- (21) Kästner, J. *Chem.—Eur. J.* **2013**, *19*, 8207–8212.
- (22) Nocedal, J. *Math. Comput.* **1980**, *35*, 773.
- (23) Liu, D. C.; Nocedal, J. *Math. Program.* **1989**, *45*, 503.
- (24) Müller, K.; Brown, L. D. *Theor. Chim. Acta* **1979**, *53*, 75.
- (25) Dellago, C.; Bolhuis, P. G.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 9236.
- (26) Powell, M. J. D. *Math. Prog.* **1971**, *26*, 1.
- (27) Bofill, J. M. *J. Comput. Chem.* **1994**, *15*, 1.
- (28) Olsen, R. A.; Kroes, G. J.; Henkelman, G.; Arnaldsson, A.; Jónsson, H. *J. Chem. Phys.* **2004**, *121*, 9776.