Reliable Transition State Searches Integrated with the Growing String Method

Paul Zimmerman*

Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States

S *Supporting Information*

ABSTRACT: The growing string method (GSM) is highly useful for locating reaction paths connecting two molecular intermediates. GSM has often been used in a two-step procedure to locate exact transition states (TS), where GSM creates a quality initial structure for a local TS search. This procedure and others like it, however, do not always converge to the desired transition state because the local search is sensitive to the quality of the initial guess. This article describes an integrated technique for simultaneous reaction path and exact transition state search. This is achieved by implementing an eigenvector following optimization algorithm in internal coordinates with Hessian update techniques. After partial convergence of the string, an exact saddle point search begins under the constraint that the maximized eigenmode of the TS node Hessian has significant overlap with the string tangent near the TS. Subsequent optimization maintains connectivity of the string to the TS as well as locks in the TS direction, all but eliminating the possibility that the local search leads to the wrong TS. To verify the robustness of this approach, reaction paths and TSs are found for a benchmark set of more than 100 elementary reactions.

Transition States from GSM
- 100+ test examples
- 100% success rate
- Average of 500 gradients required
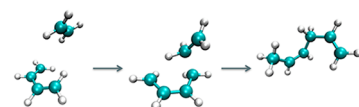
INTRODUCTION

Quantum chemistry is now sufficiently advanced that providing detailed descriptions of chemical mechanisms is almost routine. If the reactant, product, and transition state (TS) for a given elementary chemical step are known, these methods provide an incredible level of detail, specifying the kinetics as well as the electronic effects that govern the reaction.[1−7] Key to the success of quantum chemical descriptions is the systematic location and optimization of the stable intermediates and TSs in the reaction.[8−46] Intermediates are generally easy to find, being local minima with relatively simple bonding. TSs, however, remain a challenge for systematic optimization because they have significantly more complex structures: not only are TSs saddle points on the potential surface but they also are much less predictable via chemical intuition compared to stable intermediates.

Double-ended string methods[47−74] are surprisingly useful techniques for generating approximate TS structures. These methods form reaction paths (RPs) connecting two intermediates which serve as end points of the string. The reaction path is represented by interpolations between a finite number of structures along the reaction path, and thus these methods are often called "chain-of-states" methods. String optimization leads to a characterization of the potential energy profile of the reaction, which inherently has at least one local maximum corresponding to an approximate TS. Usually, a local maximum along the path is an approximate TS structure because the reaction tangent direction is also approximate. This means that even using a "climbing image" search[54] along the reaction path direction—where the highest energy node is pushed toward the local maximum—does not ultimately converge to the exact TS.

In part due to this issue, and also because optimizing a reaction path to tight convergence takes substantial computational effort, string methods are often iterated only until a reasonable TS approximation is available.[61,68,71−73] Afterward, the double-ended string methods' approximate TS serves as an initial guess for a local TS search[8−41] to find the exact TS. The dimer method[25] or partitioned rational function optimization (P-RFO) eigenvector following method[29,40] can be used for this local search. The success of these routines is known to be highly dependent on the quality of the initial guess structure.[29,68,71−73] In case of failure, the double-ended string method must be converged to tighter tolerances (possibly by adding more nodes) in order to construct a better approximate TS structure. This comes at additional computational cost and also requires another manual step to restart the searches.

These difficulties can be overcome by integrating reaction path and TS optimization routines in a double-ended string method that includes an exact local TS search algorithm. This strategy leads to reliable TS finding and reasonable assurance that the exact TS lies along the reaction path. To achieve this goal, the growing string method (GSM) is combined with an eigenvector following TS search. The new method is described in the following section, and many examples demonstrate its high reliability.

METHODS

Growing String Method in Internal Coordinates. Because the new method builds upon GSM using ICs,[74] a

brief summary of GSM in ICs is provided. In general, GSM is designed to optimize reaction paths while requiring significantly fewer gradient computations than previous string methods.[68] GSM does this by adding nodes along the reaction path only after the existing nodes are fairly well optimized. This avoids problems with poor reaction tangent directions[71,74] that can lead to node placement in high energy regions and slow reaction path convergence. During string optimization, reparameterization of the nodes along the string must occur to maintain even spacing. Convergence of GSM can be monitored by the sum of the gradient magnitudes perpendicular to the reaction tangent at each node.

GSM has recently been implemented using delocalized internal coordinates[34] constructed from a set of primitive internals defined at the reactant and product nodes.[74] The delocalized coordinates consist of a set of $3N - 6$ degrees of freedom that precisely span the nonredundant internal coordinate space. Not only does this accelerate convergence by providing an improved coordinate system for optimization (compared to Cartesians), but reaction tangents are also greatly improved, leading to higher-quality node placement and less need for subsequent optimization. The RP tangent between nodes i and j is defined as

$$U_c = \alpha_c \sum_k (q^{p,j} - q^{p,i} | U_k^i) U_k^i \tag{1}$$

where U_c is the (constrained) tangent direction, q^p are the primitive coordinates, α_c is a normalization factor, and the vectors U_k are the delocalized ICs. U_c can be projected out of the remaining ICs to form a nonredundant set spanning $3N - 7$ degrees of freedom, and the remaining vector becomes U_c.

In this coordinate system, projection of the tangential forces from the gradient at any given node is not necessary. Instead, the optimization is constrained to the $3N - 7$ dimensional space perpendicular to U_c, allowing a possibility of using approximate Hessian schemes to accelerate convergence. The constraint vector U_c can be modified to walk uphill on the apparent TS node, enabling a simple method to perform a climbing image reaction path search.[54] More details of this approach are provided elsewhere.[74]

Eigenvector Optimizer. Given a set of delocalized ICs and an appropriate Hessian, diagonalization of that Hessian in $3N - 7$ coordinates (without coordinate U_c) provides a set of eigenvectors spanning those coordinates. Optimization can then proceed by

$$\Delta q_i = \frac{-g_i}{H_{ii} + \lambda} \tag{2}$$

where q_i are the eigenvectors of the Hessian, H_{ii} are the corresponding eigenvalues, g_i is the gradient in the eigenvector basis, and λ is a scaling factor. This scheme works well given appropriate restrictions on step length and a reasonably accurate Hessian.[29,34,36,41]

Once the string is converged to a desired tolerance, a climbing image[54,74] search can move the highest energy node upward closer to the exact TS. The climbing image has been shown to greatly improve the quality of the approximate TS during string optimization.[74] Climbing image steps are taken using a steepest ascent algorithm operating on the TS node in the direction of the reaction path. Note that this does not constitute an exact TS search but will give a better approximate structure for the exact TS search when it commences.

During each string iteration, up to three optimization steps are taken on each node. A maximum of two steps is taken per node during the growth cycle. If the perpendicular gradient on any node reaches the convergence tolerance, optimization ceases on that node until the next iteration.

Hessian Construction and Update. An initial Hessian can be constructed by converting an approximate diagonal Hessian in the primitive ICs into a delocalized IC Hessian[34,35] through

$$H = U^T H^{\text{prim}} U \tag{3}$$

The initial diagonal values of H^{prim} are chosen corresponding to bonds, angles, and torsions, as needed. While this is known to be a good initial Hessian, Hessian update schemes must be employed to ensure smooth convergence. One suitable scheme, the Broyden−Fletcher−Goldfarb−Shanno (BFGS) method,[42−45] updates the Hessian as

$$\Delta H_{\text{BFGS}} = \frac{\Delta g \Delta g^t}{\Delta g^t \Delta x} - \frac{H_{i-1} \Delta x \Delta x^t H_{i-1}}{\Delta x^t H_{i-1} \Delta x} \tag{4}$$

where Δg and Δx are constructed from the current and previous gradients and coordinates.

BFGS cannot be used "as-is" in the delocalized IC basis because the ICs change after each reparameterization. However, a solution to this difficulty is to build and maintain a Hessian in the primitive internal coordinates, which are consistent throughout the optimization. The BFGS update can be applied by forming Δg and Δx in the primitive basis from the analogous vectors in the delocalized basis. After each update or after reparameterization, the Hessian in delocalized ICs can be formed using eq 3.

Exact Transition State Search. Eigenvector following TS searches can be performed after modifying eq 2 by diagonalizing the Hessian in all $3N - 6$ coordinates, then minimizing along $3N - 7$ coordinates and maximizing along one coordinate corresponding to the TS mode. In conventional searches, the TS direction is identified as the lowest eigenvalue mode of the Hessian (with a negative eigenvalue). In the context of an isolated local TS search with an exact Hessian, this procedure works well if the approximate TS structure is close to the true TS. In a reaction path optimization without exact Hessian information, however, an alternative strategy can be proposed.

The starting Hessian for the exact TS search is constructed by modifying the TS node's existing Hessian matrix in the coordinates U. Specifically, the curvature, C, along the reaction path at the TS node is approximated using the two neighboring nodes

$$C = \frac{2E_{TS-1}}{a(a + b)} - \frac{2E_{TS}}{ab} + \frac{2E_{TS+1}}{b(a + b)} \tag{5}$$

where a is the distance between the TS and the previous node and b is the distance to the following node.[75] This apparent TS direction can be added to the Hessian by applying

$$\Delta H = (C - U_c^T H U_c) U_c U_c^T \tag{6}$$

to the existing positive definite Hessian.[49,72] This transformation creates a Hessian with exactly one direction of negative curvature corresponding to the current tangent.

This Hessian can be subsequently updated during optimization steps using the Bofill update[46]

$$\Delta H_{\text{Bofill}} = \phi \Delta H_{\text{MS}} + (1 - \phi) \Delta H_{\text{PSB}} \tag{7}$$

B

dx.doi.org/10.1021/ct400319w | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

which combines the Murtagh Sargent (MS) update

$$\Delta H_{MS} = \frac{(\Delta g - H\Delta x)(\Delta g - H\Delta x)^t}{(\Delta g - H\Delta x)^t \Delta x} \tag{8}$$

with the Powell Symmetric Broyden (PSB) update

$$\Delta H_{PSB} = \frac{(\Delta g - H\Delta x)\Delta x^t + \Delta x(\Delta g - H\Delta x)^t}{\Delta x^t \Delta x}$$
$$- \frac{\Delta x^t(\Delta g - H\Delta x)\Delta x\Delta x^t}{(\Delta x^t \Delta x)^2} \tag{9}$$

using a mixing factor

$$\phi = \frac{((\Delta g - H\Delta x)^T \Delta x)^2}{|\Delta g - H\Delta x|^2 |\Delta x|^2} \tag{10}$$

This Hessian update strategy allows negative eigenvalues in the Hessian (unlike BFGS) and has previously been useful for single-ended TS searches.[29,35,41,72,73] Because reparameterization does not affect the TS node, the coordinates U for the TS node will remain applicable for the remainder of the optimization. Therefore, the Bofill update is applied directly to the Hessian in coordinates U, not the primitive coordinate Hessian.

Using the above to form and update the Hessian matrix, exact TS searches can be performed using eigenvector following (without applying constraint U_c). These searches are commenced after the string has undergone climbing image iterations and reached a given convergence threshold. While conventional eigenvector following will ascend along the lowest eigenvalue vector of the Hessian, this criterion can be modified for increased stability since the approximate reaction path direction is known in GSM. To ensure that the correct mode is followed, the eigenvector of the Hessian with the highest overlap with the string tangent at the TS is targeted. While this eigenvector must be the only one with a negative eigenvalue at a first order saddle point, multiple negative eigenvalues can appear in the Hessian during optimization. These usually lead to an ambiguity in which eigenvector to ascend, and in practice the overlap metric prevents this problem.

Multiple negative eigenvalues do not only affect the convergence of the TS mode, but also the convergence of the other $3N - 7$ modes. Therefore it is helpful to occasionally reset the Hessian when more than one negative eigenvalue is present. This is achieved by repeating the application of eqs 5 and 6, which works best when the primitive Hessian is up to date. Therefore a positive definite primitive Hessian is also maintained (using BFGS) at the TS node specifically for this purpose.

Generally, the eigenvector overlap with the reaction path tangent at the TS node is less than 1 because the reaction path tangent is not precisely an eigenvector of the Hessian (see Figure 1). This is also why a climbing image search cannot usually reach the exact TS—the reaction path tangent is an approximation to the TS mode because no Hessian information is used in its formation. However, the reaction path tangent is rarely qualitatively wrong because it lies in the direction of the potential energy reaction valley in which the string tends to converge.

Once the exact TS search commences, string optimization continues to proceed as usual. The TS node itself is not reparameterized, but equal spacing of nodes on either side of the string is still enforced. Because the maximized eigenvector
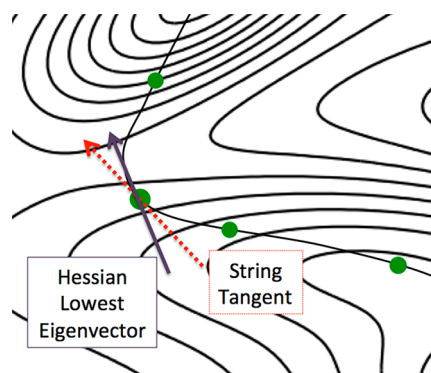

Figure 1. Qualitative comparison of string tangent with Hessian TS eigenvector on the Mueller Brown[76] potential.

of the TS node must overlap with the reaction path tangent, the TS node remains effectively constrained within the reaction path.

Summary of GSM with Exact TS Search. The overall proposed algorithm (see Figure 2) for reaction path
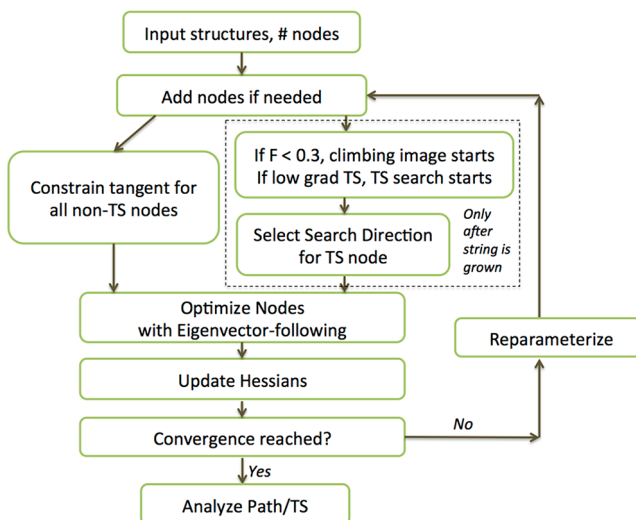

Figure 2. Process flow for growing string method with exact TS search.

optimization has some similarities to using a climbing image search in GSM with rapidly converging Hessian guided optimization. Actually, this algorithm begins with a climbing image search until reaching a specified convergence threshold and then commences the TS search. This search operates via the eigenvector following routine, which serves to pin down the exact TS using Hessian eigenvectors.

GSM requires only two input structures and the number of nodes in order to operate. The algorithm cycles through an optimization loop using the constrained eigenvector following of eq 2, updating the Hessians, checking for convergence, and finally reparameterizing the string (Figure 2). New nodes are added if the frontier nodes are within a specified gradient and the maximum number of nodes has not been reached. Once the string is grown and the total gradient falls low enough, a climbing image search and later the eigenvector following TS search begin. The eigenvector search at the TS is allowed twice as many optimization iterations per cycle (six) as the other nodes (three). Optimization on a node is halted (to limit
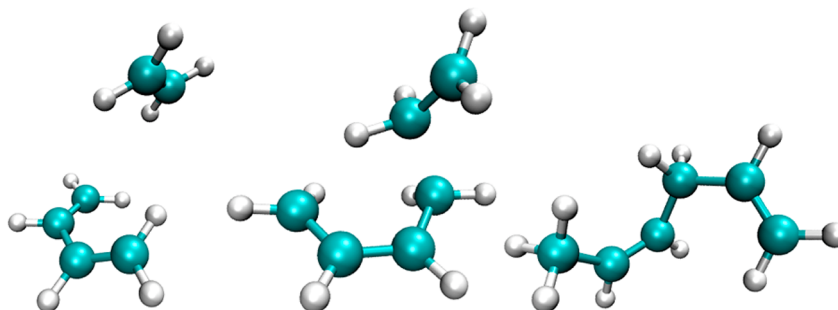
Figure 3. Ene reaction reactant, transition state, and product (from left to right).

gradient calls) until the subsequent iteration if the node reaches a nodal convergence threshold.

COMPUTATIONAL DETAILS

The B3LYP density functional[77−79] in a spin unrestricted formalism with the double-ζ, polarized 6-31G** basis set chosen for the DFT computations. The method is implemented in C++ and invokes Q-Chem 4.0[80] to provide the quantum mechanical gradients. Frequency computations on all TS structures verified that each contained the negative eigenvalue corresponding to the TS normal mode. Images of intermediates were generated in VMD.[81]

In the examples that follow, between 7 and 11 nodes including the two fixed end points are used to represent the string. While the string is growing, new nodes are added when the perpendicular gradient magnitude on the frontier node falls below 0.15 hartree/Å. The eigenvalue shift parameter λ is chosen such that the shifted eigenvalues are all positive, except for the exact TS search, where λ shifts only the non-TS eigenvalues. A maximum optimization step size of 0.1 is applied for all test cases. Convergence of the reaction path is considered complete when the TS node has a small root mean squared (RMS) gradient (<0.0005 hartree/Å). The same RMS gradient threshold is used to temporarily cease node optimization during each string iteration. The climbing image search is initiated after the sum of the perpendicular gradient magnitudes over all nodes, F, is converged to $F < 0.3$. The exact TS search is initiated after $F < 0.1$, or when the highest energy node has a RMS gradient below twice the nodal convergence criteria and $F < 0.2$.

RESULTS AND DISCUSSION

Ene Reaction Convergence. To gain insight into the convergence behavior of the new method, one test case, an ene reaction between *cis*-butadiene and ethylene, will be investigated in detail before proceeding to an extended test set. This reaction proceeds through simultaneous H transfer with C−C bond formation, shown in Figure 3. GSM-IC with eigenvector following is able to converge the reaction path and exact TS with a barrier of 41 kcal/mol using 365 gradient computations. In other methods where the TS search is performed separately from the string, the exact Hessian would need to be computed to confirm the accuracy of the TS structure. With GSM, however, if the TS exists as a maximum along the reaction path with a converged gradient and TS eigenvector overlapping with the reaction path tangent, this check is not as necessary (specifically, a stationary point of this type must be a saddle point of at least first order). Regardless, the exact Hessian

computed at this structure reveals a single negative eigenvalue corresponding to the desired TS direction.

The convergence behavior of the reaction path and TS can be compared to other GSM methods using ICs. For comparison with GSM in Cartesian coordinates, see ref 74, where it was found that GSM in ICs is more rapidly converging than in other coordinate systems. Shown in Figure 4 is the sum
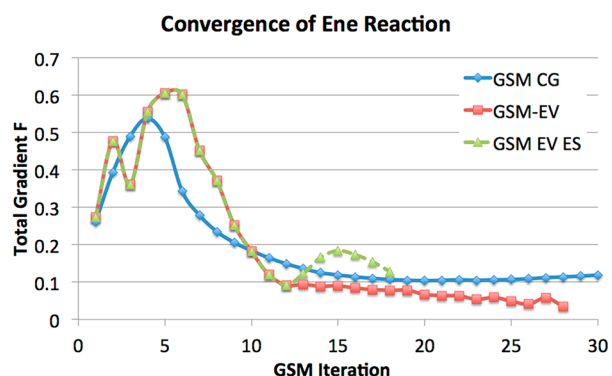

Figure 4. Convergence behavior of GSM in ICs comparing GSM using conjugate gradient to eigenvector following with and without exact TS search. All three methods employ climbing image searches after $F < 0.3$. The exact TS search begins for GSM-IC-EV after $F = 0.1$.

of the perpendicular gradient magnitudes, F, as a function of GSM iteration. Complementary to this is Figure 5, which shows the energy of the TS node as the optimization proceeds. In Figures 4 and 5, GSM in ICs with the conjugate gradient optimizer[74] (GSM-CG) and the optimizer of eq 2 (GSM-EV)
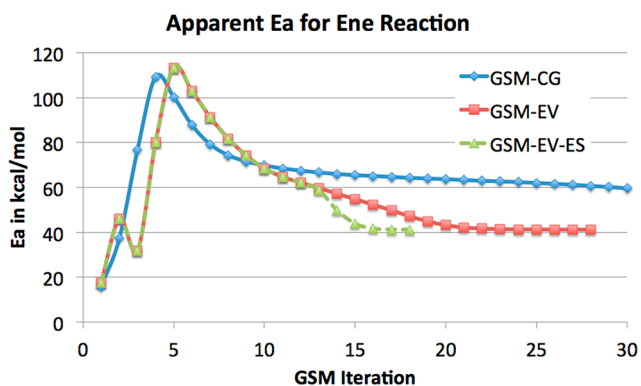

Figure 5. TS node energy as a function of optimization iteration comparing GSM using conjugate gradient (GSM-CG) to eigenvector following with (GSM-EV-ES) and without (GSM-EV) exact TS search. All three methods employ climbing image searches.

D

dx.doi.org/10.1021/ct400319w | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

are compared to the second method employing the exact TS search (GSM-EV-ES). All three use climbing image, and the node specific optimization stop criterion (see Methods section) is turned off for more straightforward comparisons. The exact TS search initiates after the string is converged to $F < 0.1$.

The first item of note from Figure 4 is that the conjugate gradient optimizer performs better than the eigenvector optimizer in early stages of optimization. This can occur because conjugate gradient performs well in high gradient regions and does not rely on Hessian information. The eigenvector optimizer is slower at first when the initial approximate Hessian has not yet been made accurate by the update schemes. This is also evidenced by the fact that GSM-EV starts to converge faster than GSM-CG soon after the string is done growing, and GSM-EV reaches $F = 0.1$ faster than GSM-CG.

The exact TS search for GSM-EV-ES begins just after iteration 12 by removing constraint U_c and adding the TS curvature to the Hessian via eqs 5 and 6. The total gradient F increases with the next few subsequent iterations as the TS node moves away from the maximum on the reaction path tangent, dragging the remainder of the string to a slightly different pathway. This occurs because the Hessian update modifies the direction of negative curvature during optimization, turning the TS eigenvector away from the reaction path tangent. During this time, the vector overlap between the Hessian TS vector and the reaction path tangent drops to as low as 78%, and eventually rises to 97% at the final GSM iteration. At iteration 18, GSM-EV-ES has reached the exact TS and optimization is complete.

Although the exact TS search can appear to make convergence slower (by increasing F), in reality the entire string bends quickly toward the true reaction path when the TS node moves to the exact TS. (If optimization were continued after the TS was found, the remainder of the string would converge rapidly because the TS node is fixed.) In contrast, GSM-CG with climbing image is not converged within the 30 iterations shown in the figures, and GSM-EV's TS node converges at 28 iterations (see Figure 4). GSM-EV converging to the TS via climbing image therefore represents a case where the reaction path tangent can become close to the TS eigenvector, although this does not occur in every reaction. In this example reaction, it is apparent that GSM-EV-ES is superior because it is able to converge to the exact TS faster than GSM-EV. This is evidenced by the quick asymptote of GSM-EV-ES to the exact TS energy in Figure 5, whereas GSM-EV requires 10 additional GSM iterations to reach convergence. Overall, these results show that GSM-EV-ES can find exact TSs and approximate RPs faster than previous optimizers (conjugate gradient) and approximate TS locating techniques (climbing image).

Test Set. A test library was created in order to verify the reliability of the proposed method. The chemical reactions included in the set include a few examples from previous TS finding articles,[41,82,83] reactions of the hydrogen storage material ammonia-borane and NH_2BH_2,[84−87] and a variety of other species relevant to organic chemistry.[88−91] Most of these examples were created with the assistance of an automated elementary step generator, which proposes hypothetical reaction steps starting from a single structure.[82] From these hypothetical steps, reactions that were confirmed to be single elementary steps were taken and added to the test set.

In this set, the smallest system has five atoms, while the largest has 53. The elementary steps include a wide variety of reactions, spanning low to high barrier heights (3 to 127 kcal/mol). Summaries of these reactant species are given in Table 1

Table 1. Elementary Step Test Cases for the New Growing String Method

reactant	# paths	path descriptions
		set 1
2 NH_2BH_2	7	B−H, N−H additions, transfers
NH_2BH_2 + NH_3BH_3	9	NH_3, BH_3, N−H, B−H additions, transfers, H_2 formation
propene	2	H transfer, cyclopropane formation
acetaldehyde	1	H transfer
formaldehyde + NH_3	4	N−H, C−H transfer, N−H addition, H_2 elimination
H_2 + SiH_2	1	H_2 addition
ethane	1	H_2 elimination
formaldehyde + vinyl alcohol	13	C−C bond formations, H, O transfers, O−H addition
ethylene + $HPMe_2$	1	ethylene insertion
SiO_2Cl_2	2	Cl transfer, Cl_2 elimination
formaldehyde + methanol	7	ether formation, O−H addition, H transfers, H_2 formation
2-butene-2-ol	5	H transfer, three member ring formation, dehydration
2-butene-2-ol + water	8	O−H addition, H transfers, OH transfer, H_2 formation
ethyl trifluoroacetate + NH_3	3	N−H additions, ethyl transfer
fluorinated ethylene oxide	3	CF_3 transfer, F transfers
vinyl-methyl amine	5	C−H transfer, ring formation, methane elimination
		set 2
taxadiene carbocation	3	proton transfers
alanine dipeptide	1	dihedral rotation
cis-butadiene + ethylene	13	ring formation, ene reactions, H transfers, C−C bond exchange
1,4 methyl-cyclobutane	1	ring-opening
2-butene + peracetic acid	4	OH transfer, C−H addition, epoxide formation
2-butene-1-ol + $SiOCl_2$	5	O transfers, HCl eliminations, epoxide formation
ethylene oxide + PMe_3	1	epoxide addition
isobutene + water	1	hydration
allyl phenyl ether	4	C_3H_5 rotation, epoxide formation, three member ring formation

along with brief descriptions of the reactions. More information, including geometries and activation barriers, is provided in the Supporting Information. These reactions span from almost trivial reactions that proceed over a single coordinate to complex multiatom rearrangements with substantial curvatures over the reaction path. The present test set is larger by an order of magnitude compared to previous string method test sets which have typically contained less than 10 total elementary reactions.[47−74]

Benchmarks. Three factors are of prime importance for the success of the present method: the difficulty of the TSs, consistency in locating the desired TSs, and the number of gradient computations required for convergence. Having a large

E

dx.doi.org/10.1021/ct400319w | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

test set stresses all three points and ensures that a string method is stretched to its limits.

Using a string with 11 nodes, GSM-EV-ES is able to locate each reaction path and TS in Table 1. This constitutes over 100 reactions, which required on average 500 gradients per reaction to converge. To test the dependence on the number of nodes and other optimization parameters, additional benchmarking was performed.

A subset of 72 reactions in Table 1 was studied in detail for this purpose (set 1). Set 1 was chosen due to the relatively smaller size of the molecules, facilitating comparisons to be made between methods. Table 2 shows the different operating

Table 2. Operating Conditions for the New Growing String Method

name	# nodes	max iterations	climbing image?	overlap criterion?
7 nodes	7	80	yes	yes
9 nodes	9	80	yes	yes
11 nodes	11	80	yes	yes
11 nodes no climb	11	80	no	yes
11 nodes no overlap	11	80	yes	no

conditions for GSM: a range of 7, 9, or 11 nodes, and GSM with or without climbing image or the overlap criterion. Reducing the number of nodes should reduce the number of gradient calls on average, but the degree of reaction path accuracy should decrease and there will be cases that fail to converge. Turning off climbing image before the exact TS search will reduce the initial quality of the TS guess, also decreasing performance. Similarly, the overlap criterion prevents wandering away from the exact TS during optimization, so removing this criterion will lead to increased numbers of optimization steps or failure of the TS search.

After running the reduced test set through the conditions of Table 2, all of the possibilities of the previous paragraph were realized. Specifically, Figure 6 shows the success rate alongside
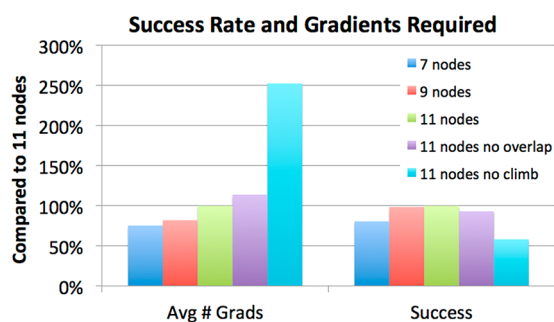

Figure 6. Success rate and number of gradients required for GSM-EV-ES with different optimization conditions on test cases in set 1 of Table 1. On average, 468 gradients were required to converge with 11 nodes.

the number of gradient calls required. The number of gradients is normalized to that of GSM-EV-ES with 11 nodes, which required 468 gradients on average to converge set 1. Note that optimization is halted after 80 string iterations, so failed runs do not excessively contribute to the statistics. Notably, while both the overlap criterion and climbing image are important for successful convergence, the climbing image is significantly more important. Removing the overlap criterion decreases the

number of successful runs to 93%, but removing climbing image results in a success rate of only 58%. Furthermore, because so few runs completed successfully without climbing image (i.e., before reaching 80 GSM iterations), the average number of gradient calls required for GSM without climbing image jumps to almost 1,200. These results suggest that the climbing image is vital to starting the eigenvector following TS search in a region where it can more easily converge to the true TS.

This situation is also reflected in Figure 7, which compares the total gradient F, RMS gradient, and final overlap values.
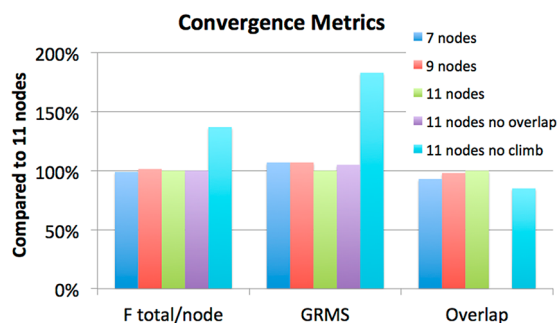

Figure 7. Convergence measures for GSM with different optimization conditions, normalized to the 11 node GSM-EV-ES with overlap and climbing features, on test cases in set 1 of Table 1.

Again, climbing image is very helpful for string convergence, increasing the final overlap and decreasing both F and the RMS gradient. The RMS gradient and overlap are similar for nine and 11 nodes, suggesting that the number of gradient computations can be reduced by using just nine nodes in many cases (only one failed TS search in set 1 with nine nodes). Going to seven nodes, however, shows lower vector overlap, which is consistent with a significantly reduced success rate (81%). Therefore seven node operation cannot be recommended with this method, where the smaller number of nodes leads to less reliable convergence of difficult RPs. Because 11 nodes were sufficient to converge all test cases, there is no particular need to increase their number. The node count could be increased, however, if a specific reaction was deemed to be particularly complex.

CONCLUSIONS

TS location has long been a substantial challenge, leading to the development of many algorithms over the past few decades. TS searches are difficult to perform due to the complexity of chemical rearrangements and high computational cost of quantum mechanical methods, so increasing reliability and decreasing cost have always been desirable targets. Considering these criteria and the 100+ examples tested herein, GSM-EV-ES performs well and is able to find TSs in all cases using just 11 nodes. At an average cost of 500 gradients per TS/reaction path combination (and no exact Hessian required for the TS search), the method can be efficiently used in a variety of chemical systems. The two fastest converging examples were alanine dipeptide and phenyl ether torsional rotations, requiring 80 and 114 gradients, respectively. These were particularly simple cases, being the only two in the test set that did not involve bond breaking or forming. The two slowest converging test cases were H exchange between NH_3 and H_2CO, and epoxide addition to PMe_3, requiring 1068 and 1520 gradients,

respectively. These latter two cases were the only two that required more than 1000 gradients to converge and likely were difficult due to the radical character of the reactions. Overall, GSM-EV-ES works well for single elementary step processes, but in principle the method could be extended to locate multiple exact TSs along a path.

The test cases show that using climbing image is effective at approaching the true TS consistently and that an eigenvector search using approximate Hessians works well when the search targets the Hessian eigenvectors with highest overlap to the reaction path tangent. Removal of either of these strategies results in significantly fewer converged TSs, indicating the necessity of each of these strategies. Additional developments could include extending the method to large systems using QM/MM.[92−97]

An additional key benefit of the new GSM is the combination of high reliability with low degree of user labor: only the reactant and product structures must be prepared as input, and the exact TS along the reaction path is found without having to start a second computation. This technique will greatly benefit automatic elementary step generation, for instance using a method recently developed by the author, where GSM can provide the reaction paths for pairs of reactant/product structures.[82] Overall, the new version of GSM can be expected to be useful for standard TS searches whenever the reactant and product of a reaction are known.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Cartesian geometries of all reactants, products, and transition states. This information is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: paulzim@umich.edu.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Ziegler, A. T. *Chem. Rev.* **1991**, *91* (5), 651−667.
(2) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100* (31), 12974−12980.
(3) Parr, R. G.; Yang, W. *J. Am. Chem. Soc.* **1984**, *106*, 4049−4050.
(4) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1* (3), 415−432.
(5) Cramer, C. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2009**, *11* (46), 10757−10816.
(6) Bell, A. T.; Head-Gordon, M. *Annu. Rev. Chem. Biomol. Eng.* **2011**, *2*, 453−477.
(7) Keil, F. J. *Top. Curr. Chem.* **2012**, *307*, 69−108.
(8) Cerjan, C. J.; Miller, W. H. *J. Chem. Phys.* **1981**, *75*, 2800.
(9) Simons, J.; Jorgensen, P.; Taylor, H.; Ozment, J. *J. Phys. Chem.* **1983**, *87*, 2745−2753.
(10) Cancès, E.; Legoll, F.; Marinica, M. C.; Minoukadeh, K.; Willaime, F. *J. Chem. Phys.* **2009**, *130*, 114711.
(11) Poppinger, D. *Chem. Phys. Lett.* **1975**, *35*, 550−554.
(12) Peters, B.; Liang, W.-Z.; Bell, A. T.; Chakraborty, A. *J. Chem. Phys.* **2003**, *118* (21), 9533.
(13) Laio, M.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562−12566.
(14) Ianuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90* (23), 238302.
(15) Ensing, B.; de Vivo, M.; Liu, Z.; Moore, P.; Klein, M. L. *Acc. Chem. Res.* **2006**, *39*, 73−81.
(16) Mueller, E. M.; de Meijere, A.; Grubmueller, H. *J. Chem. Phys.* **2002**, *116* (2), 897.
(17) Samanta, A.; E, W. *J. Chem. Phys.* **2012**, *136*, 124104.
(18) Burger, S. K.; Ayers, P. W. *J. Chem. Phys.* **2010**, *132*, 234110.
(19) Chen, M.; Cuendet, M. A.; Tuckerman, M. E. *J. Chem. Phys.* **2012**, *137*, 24102.
(20) Maeda, S.; Ohno, K.; Morokuma, K. *J. Chem. Theory Comput.* **2009**, *5*, 2734−2743.
(21) Maeda, S.; Morokuma, K. *J. Chem. Theory Comput.* **2011**, *7*, 2335−2345.
(22) Maeda, S.; Abe, E.; Hatanka, M.; Taketsugu, T.; Morokuma, K. *J. Chem. Theory Comput.* **2012**, *8* (12), 5058−5063.
(23) Schlegel, H. B. *J. Comput. Chem.* **1982**, *3*, 214−218.
(24) Henkelman, G.; Jonsson, H. *J. Chem. Phys.* **1999**, *111*, 7010.
(25) Heyden, A.; Bell, A. T.; Keil, F. J. *J. Chem. Phys.* **2005**, *123*, 224101.
(26) Baker, J. *J. Comput. Chem.* **1986**, *7*, 385−395.
(27) Wales, D. J. *J. Chem. Soc., Faraday Trans.* **1992**, *88*, 653−657.
(28) del Campo, J. M.; Koster, A. M. *J. Chem. Phys.* **2008**, *129*, 024107.
(29) Schlegel, H. B. *WIREs Comput. Mol. Sci.* **2011**, *1*, 790−809.
(30) Schlegel, H. B. *J. Comput. Chem.* **2003**, *24*, 1514−1527.
(31) May, J. W.; Lehner, J. D.; Frisch, M. J.; Li, X. S. *J. Chem. Theory Comput.* **2012**, *8* (12), 5175−5179.
(32) Li, X. S.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2* (3), 835−839.
(33) Liang, W.; Wang, H.; Hung, J.; Li, X. S.; Frisch, M. J. *J. Chem Theory Comput.* **2010**, *6*, 2034−2039.
(34) Baker, J.; Kessi, A.; Delley, B. *J. Chem. Phys.* **1996**, *105*, 192.
(35) Bakken, V.; Helgaker, T. *J. Chem. Phys.* **2002**, *117* (2), 9160.
(36) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177−2186.
(37) Baker, J. *J. Comput. Chem.* **1993**, *14* (9), 1085−1100.
(38) Pulay, P.; Fogarasi, G. *J. Chem. Phys.* **1992**, *96*, 2856−2860.
(39) Paizs, B.; Baker, J.; Suhai, S.; Pulay, P. *J. Chem. Phys.* **2000**, *133*, 6566−6572.
(40) Banerjee, A.; Adams, N.; Simons, J.; Shepard, R. *J. Phys. Chem.* **1985**, *89*, 52−57.
(41) Baker, J.; Chan, F. *J. Comput. Chem.* **1996**, *17* (7), 888−904.
(42) Broyden, C. G. *J. Inst. Math. Appl.* **1970**, *6*, 76−90.
(43) Fletcher, R. *Comput. J.* **1970**, *13*, 317−322.
(44) Goldfarb, D. *Math. Comput.* **1970**, *24*, 23−26.
(45) Shanno, D. F. *Math. Comput.* **1970**, *24*, 647−657.
(46) Bofill, J. M. *J. Comput. Chem.* **1994**, *15*, 1−11.
(47) Granot, R.; Baer, R. A. *J. Chem. Phys.* **2008**, *128*, 184111.
(48) Ghasemi, S. A.; Goedecker, S. *J. Chem. Phys.* **2011**, *135*, 014108.
(49) Ayala, P. Y.; Schlegel, H. B. *J. Chem. Phys.* **1997**, *107*, 375.
(50) Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **1996**, *17*, 49−56.
(51) Peng, C.; Schlegel, H. B. *Isr. J. Chem.* **1994**, *33*, 449−454.
(52) Mills, G.; Jonsson, H. *Phys. Rev. Lett.* **1994**, *72*, 1124.
(53) Henkelman, G.; Jonsson, H. *J. Chem. Phys.* **2000**, *113*, 9978.
(54) Henkelman, G.; Uberuaga, B. P.; Jonsson, H. *J. Chem. Phys.* **2000**, *113*, 9901.
(55) Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 2082.
(56) Chu, J.; Trout, B.; Brooks, B. A. *J. Chem. Phys.* **2003**, *119*, 12708.
(57) Sheppard, D.; Terrell, R.; Henkelman, G. *J. Chem. Phys.* **2008**, *128*, 134106.
(58) E, W.; Ren, W.; Vanden-Eijnden, E. *Phys. Rev. B* **2002**, *66*, 052301.
(59) E, W.; Ren, W.; Vanden-Eijnden, E. *J. Phys. Chem. B* **2005**, *109*, 6688−6693.
(60) Ren, W.; Vanden-Eijnden, E. *J. Chem. Phys.* **2007**, *126*, 164103.
(61) Burger, S. K.; Yang, W. *J. Chem. Phys.* **2006**, *24*, 054109.
(62) Burger, S. K.; Yang, W. *J. Chem. Phys.* **2007**, *127*, 164107.
(63) Chaffey-Millar, H.; Nikodem, A.; Matveev, A. V.; Krüger, S.; Rösch, N. *J. Chem. Theory Comput.* **2012**, *8*, 777−786.
(64) Jung, J.; Re, S.; Ten-no, S. *J. Chem. Phys.* **2013**, *138*, 044106.

(65) Liu, Y.; Burger, S. K.; Ayers, P. W. *J. Math. Chem.* **2011**, *49*, 1915−1927.

(66) Tao, P.; Hodoscek, M.; Larkin, J. D.; Shao, Y.; Brooks, B. R. *J. Chem. Theory Comput.* **2012**, *8*, 5035−5051.

(67) Klimes, J.; Bowler, D. R.; Michaelides, A. *J. Phys.: Condens: Matter* **2010**, *22*, 074203.

(68) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. *J. Chem. Phys.* **2004**, *120*, 7877.

(69) Goodrow, A.; Bell, A. T.; Head-Gordon, M. *J. Chem. Phys.* **2008**, *129*, 174109.

(70) Quapp, W. A. *J. Chem. Phys.* **2005**, *122*, 174106.

(71) Behn, A.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. *J. Chem. Theory Comput.* **2011**, *7* (12), 4019−4025.

(72) Sharada, S.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. *J. Chem. Theory Comput.* **2012**, *8* (12), 5166−5174.

(73) Behn, A.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. *J. Chem. Phys.* **2011**, *135*, 224108.

(74) Zimmerman, P. M. *J. Chem. Phys.* **2013**, *138*, 184102.

(75) Anoshkina, E. V.; Belyaev, A. G.; Seidel, H.-P. *Proceedings of Vision, Modeling, and Visualization (VMV)*, Erlangen, Germany, 2002.

(76) Muller, K.; Brown, L. D. *Theor. Chim. Acta* **1979**, *53*, 75−93.

(77) Becke, A. D. *Phys. Rev. A* **1998**, *38*, 3098.

(78) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(79) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(80) Shao, Y.; Molnar, L. F.; et al. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172−3191.

(81) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.

(82) Zimmerman, P. M. *J. Comput. Chem.* **2013**, *34*, 1385−1392.

(83) Perczel, A.; Farkas, O.; Jakli, I.; Topol, I. A.; Csizmadia, I. G. *J. Comput. Chem.* **2003**, *24*, 1026−1042.

(84) Zimmerman, P. M.; Paul, A.; Zhang, Z.; Musgrave, C. B. *Inorg. Chem.* **2009**, *48*, 1069−1081.

(85) Zimmerman, P. M.; Zhang, Z.; Musgrave, C. B. *Inorg. Chem.* **2010**, *49* (19), 8724−8728.

(86) Roy, L.; Zimmerman, P. M.; Paul, A. *Chem.—Eur. J.* **2011**, *17* (2), 435−439.

(87) Zimmerman, P. M.; Zhang, Z.; Musgrave, C. B. *J. Phys. Chem. Lett.* **2011**, *2*, 276−281.

(88) Diels, O.; Alder, L. *Liebigs Ann. Chem.* **1928**, *460*, 98−122.

(89) Tietze, L. F.; Kettschau, G. *Top. Curr. Chem.* **1997**, *189*, 1−167.

(90) Hong, Y. J.; Tantillo, D. J. *J. Am. Chem. Soc.* **2011**, *133*, 18249−18256.

(91) Koeksal, M.; Jin, Y.; Coates, R. M.; Croteau, R.; Christianson, D. W. *Nature* **2011**, *469*, 116−120.

(92) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760−769.

(93) Vreven, T.; Frisch, M. J.; Farkas, O.; Schlegel, H. B.; Morokuma, K. *Mol. Phys.* **2006**, *104*, 701−714.

(94) Zimmerman, P. M.; Head-Gordon, M.; Bell, A. T. *J. Chem. Theory Comput.* **2011**, *7*, 1695−1703.

(95) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185−199.

(96) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580−10594.

(97) de Vries, A. H.; Sherwood, P.; Collins, S. J.; Rigby, A. M.; Rigutto, A.; Kramer, G. J. *J. Phys. Chem. B* **1999**, *103*, 6133−6141.