High-index saddle dynamics for the automated mapping of reaction routes

Jianyuan Yin [1] **Vedant Bahadur** [2] **Zekun Shi** [3 4] **Vitalii Sagan** [2] **Victor Posligua** [2] **Qianxiao Li** [5 2] **Stephen Gregory Dale** [2 6]

[1] *School of Mathematical Sciences, Laboratory of Mathematics and Complex Systems, Ministry of Education, Beijing Normal University, Beijing 100875, China* [2] *Institute for Functional Intelligent Materials, National University of Singapore* [3] *SEA AI Lab, Singapore* [4] *School of Computing, National University of Singapore* [5] *Department of Mathematics, National University of Singapore* [6] *Department of Materials Science and Engineering, National University of Singapore*. Correspondence to: Qianxiao Li qianxiao@nus.edu.sg, Stephen Gregory Dale sdale@nus.edu.sg.

1. Introduction

Modern materials discovery AI techniques typically target the prediction of a final product with desirable properties.[1, 2] This approach disregards the synthesisability of the desired materials, still leaving significant open questions about how to make the desired materials before their properties can be experimentally tested and verified.

This work seeks as alternative route to training AI techniques by generating data focussed on the possible reaction paths of a given chemical system. We introduce the Molecular High-index Saddle Dynamics (MHiSD) method, based on previous non-molecular methods.[3, 4] This approach is intended to start from a local energy minimum of the potential energy surface (PES) and searching along the paths of slightest gradient ascent to find the closest neighbouring stationary states. By repeating this process iteratively, and including down-ward searches from higher index stationary states, full reaction networks can be generated. In principle this will provide multiple potential reaction paths towards the desired starting molecule or system by following the minimum energy path(\s). An additional wealth of significant information is generated for AI training in the form of the full network.

2. Methods

The MHiSD for index-k saddle points (k-MHiSD) on the potential energy surface has the form of:

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = -\left(I - \displaystyle\sum_{j=1}^{k} 2\boldsymbol{v}_j\boldsymbol{v}_j^\top\right)\nabla E(\boldsymbol{x}), \\ \boldsymbol{v}_i =_{\boldsymbol{v}\in\mathcal{V}_i(\boldsymbol{x},\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{i-1})} \langle \boldsymbol{v}, \nabla^2 E(\boldsymbol{x})\boldsymbol{v}\rangle, \\ \mathcal{V}_i(\boldsymbol{x},\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{i-1}) = \{\boldsymbol{v}_i \in N(\boldsymbol{x}) | \\ \qquad\qquad \langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle = \delta_{ij}, j = 1,\cdots,i, \quad i = 1,\cdots,k. \end{cases} \tag{1}$$

Note that $\boldsymbol{v}_i^\top\boldsymbol{v}_j$ represents the inner product $\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle$. The directions $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k$ defined in (1) represents the orthonormal eigenvectors corresponding to the smallest k eigenvalues of the operator $\mathcal{P}_{\boldsymbol{x}}\nabla^2 E(\boldsymbol{x})\mathcal{P}_{\boldsymbol{x}}$ on $N(\boldsymbol{x})$, where $\mathcal{P}_{\boldsymbol{x}} : \mathbb{R}^{3n} \to \mathbb{R}^{3n}$ denotes the orthogonal projection operator onto the normal space $N(\boldsymbol{x})$. The operator $\mathcal{P}_{\boldsymbol{x}}\nabla^2 E(\boldsymbol{x})\mathcal{P}_{\boldsymbol{x}}$ represents the projected Hessian onto the normal space $N(\boldsymbol{x})$, and includes the curvature information of the energy on the normal space [5]. Physically, these orthogonal directions corresponds to the lowest curvature of the potential energy surface by minimizing the Rayleigh quotients. The dynamics of $\boldsymbol{x}$ in (1) with an initial condition $\boldsymbol{x}(0) \in \mathbb{R}^{3n}$ represents an modified gradient flow, which is gradient ascent on the subspace spanned by $\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k\}$, and gradient descent on the subspace orthogonal to these vectors. Once this dynamics converges, the system locates an index-k saddle point on the potential energy surface. Note that the gradient flow of E is equivalent to the MHiSD of $k = 0$ case, so MHiSD can be regarded as a generalization of gradient flow to saddle points.

We obtain energy, gradient and hessian values using the psi4 computational package.[6] In principle any electronic structure program that outputs these values can be used. An open source implementation of the k-HiSD program will be released closer to the AI4X conference date.

3. Results

To demonstrate the efficacy of this method we have reproduced the Reaction Route Map (RRM) starting from methanoic acid produced by Maeda *et al.*[7] with our version shown in Figure 1. Note that Figure 1 is limited to only $k = 0, 1$ order saddle points, which results in the discovery of approximately 60 stationary points. As part of our k-MHiSD search we allow discovery of up to $k = 4$ saddle points, resulting in the discovery of over 300 saddle points.

4. Conclusions

This talk will discuss further details of the k-MHiSD method, the most recent data generated with this method, and the directions of future applications.

Acknowledgments

This research is supported by the Ministry of Education, Singapore, under its Research Centre of Excellence award to the Institute for Functional Intelligent Materials Project No. EDUNC-33-18-279-V12. This research is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG3-RP-2022-028).

References

[1] Amil Merchant, Simon Batzner, Samuel S Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, 2023.

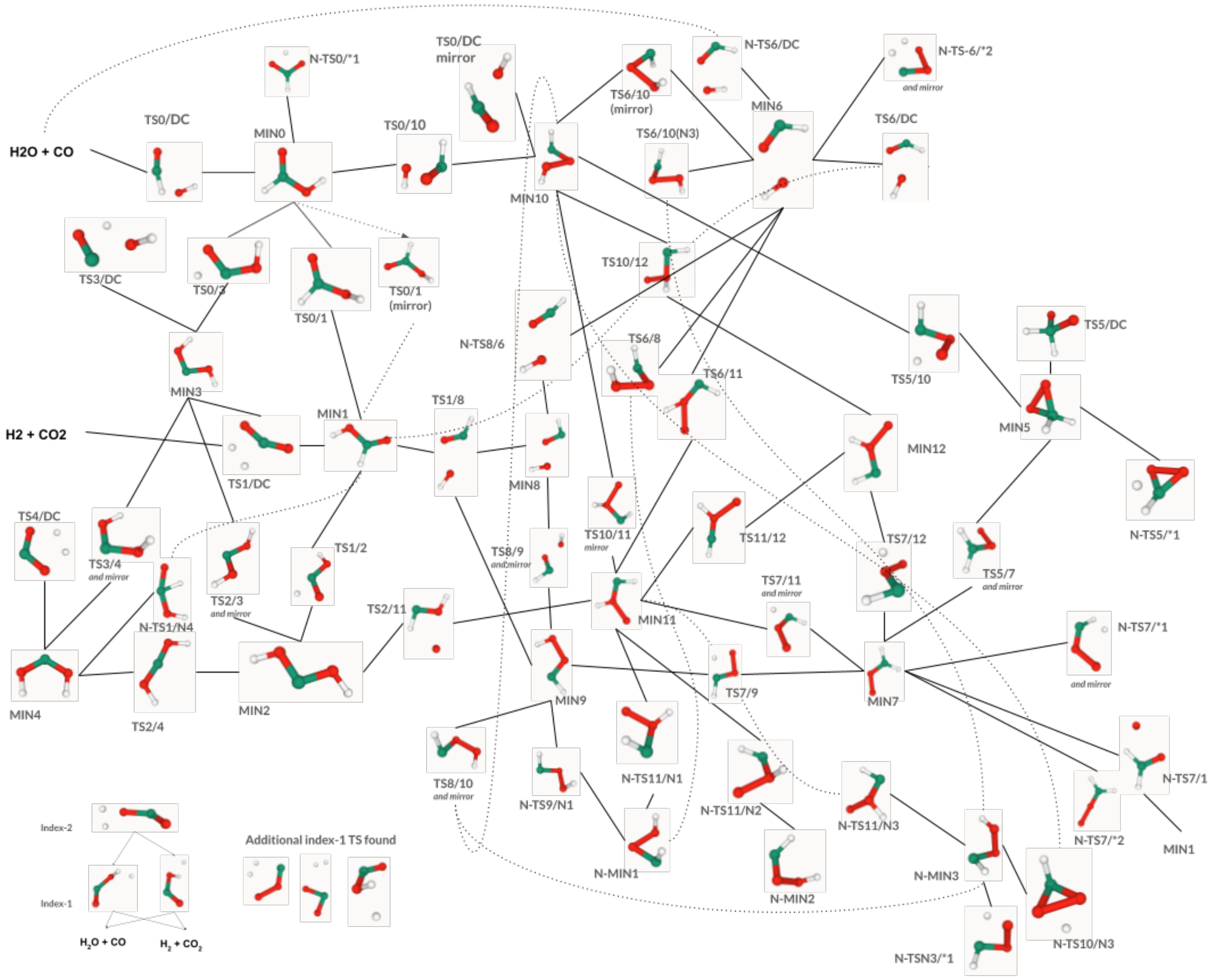

Fig. 1: The RRM of methanoic acid as produced by our k-MHiSD search method. Only order 0 and 1 saddle points are shown and arranged to best match the work of Maeda *et al.*[7].

[2] Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Sasha Shysheya, Jonathan Crabbé, Lixin Sun, Jake Smith, et al. Mattergen: a generative model for inorganic materials design. *arXiv preprint arXiv:2312.03687*, 2023.

[3] Jianyuan Yin, Lei Zhang, and Pingwen Zhang. High-index optimization-based shrinking dimer method for finding high-index saddle points. *SIAM J. Sci. Comput.*, 41(6):A3576–A3595, 2019.

[4] Jianyuan Yin, Zhen Huang, and Lei Zhang. Constrained high-index saddle dynamics for the solution landscape with equality constraints. *J. Sci. Comput.*, 91(2):62, 2022.

[5] Wolfgang Quapp and Josep Maria Bofill. Locating saddle points of any index on potential energy surfaces by the generalized gentlest ascent dynamics. *Theor. Chem. Acc.*, 133(8):1510, 2014.

[6] Justin M Turney, Andrew C Simmonett, Robert M Parrish, Edward G Hohenstein, Francesco A Evangelista, Justin T Fermann, Benjamin J Mintz, Lori A Burns, Jeremiah J Wilke, Micah L Abrams, et al. Psi4: an open-source ab initio electronic structure program. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(4):556–565, 2012.

[7] Satoshi Maeda, Tetsuya Taketsugu, Keiji Morokuma, and Koichi Ohno. Anharmonic downward distortion following for automated exploration of quantum chemical potential energy surfaces. *Bull. Chem. Soc. Jpn.*, 87(12):1315–1334, 2014.