nimCSO: A Nim package for Compositional Space Optimization

Adam M. Krajewski [1][¶], **Arindam Debnath** [1], **Wesley F. Reinhart** [1,2], **Allison M. Beese** [1], **and Zi-Kui Liu** [1]

1 Department of Materials Science and Engineering, The Pennsylvania State University, USA **2** Institute for Computational and Data Sciences, The Pennsylvania State University, USA ¶ Corresponding author

Summary

`nimCSO` is a high-performance, low-level tool for selecting components (dimensions) in compositional spaces which optimize the data availability for applications such as machine learning, which is a combinatorially hard problem for complex compositions existing in highly dimensional spaces due to the interdependency of components being present. Such spaces are often encountered in materials science, where datasets on Compositionally Complex Materials (CCMs) often span 20-40 chemical elements, while each data point contains several of them.

This tool employs a set of methods, ranging from (1) brute-force search through (2) genetic algorithms to (3) a newly designed search method. They use custom data structures and procedures written in Nim language, which are compile-time optimized for the specific problem statement and dataset pair, which allows nimCSO to run faster and use 1-2 orders of magnitude less memory than general-purpose data structures. All configuration is done with a simple human-readable config file, allowing easy modification of the search method and its parameters.

Statement of Need

`nimCSO` is an interdisciplinary tool applicable to any field where data is composed of a large number of independent components and their interaction is of interest in a modeling effort, ranging from social sciences like economics, through medicine where drug interactions can have a large impact on the treatment, to chemistry and materials science, where the composition and processing history are critical to resulting properties. The latter has been the root motivation for the development of `nimCSO` within the **ULTERA Project** (ultera.org) carried under the **US DOE ARPA-E ULTIMATE** program which aims to develop a new generation of ultra-high temperature materials for aerospace applications, through generative machine learning models (Debnath et al., 2021) driving thermodynamic modelling and experimentation (Li et al., 2023).

One of the most promising materials for these applications are Compositionally Complex Materials (CCMs), and their matal-focused subset of Refractory High Entropy Alloys (RHEAs), which are quickly growing since first proposed by (Cantor et al., 2004) and (Yeh et al., 2004). Contrary to most traditional alloys, they contain a large number of chemical elements (typically 4-9) in similar proportions, in hope to thermodynamically stabilize the material by increasing its configurational entropy ($\Delta S_{conf} = \Sigma_i^N x_i \ln x_i$ for ideal mixing of N elements with fractions x_i), what encourages sampling a large palette of chemical elements. The resulting compositional spaces are both extremely vast and challenging to explore in terms of possible changes (Krajewski et al., 2024); thus, it becomes critical to answer the question like *"Which combination of 15 elements out of 60 in the dataset will result in the largest dataset?"* which has $\binom{60}{15}$ or 53 trillion possible solutions.
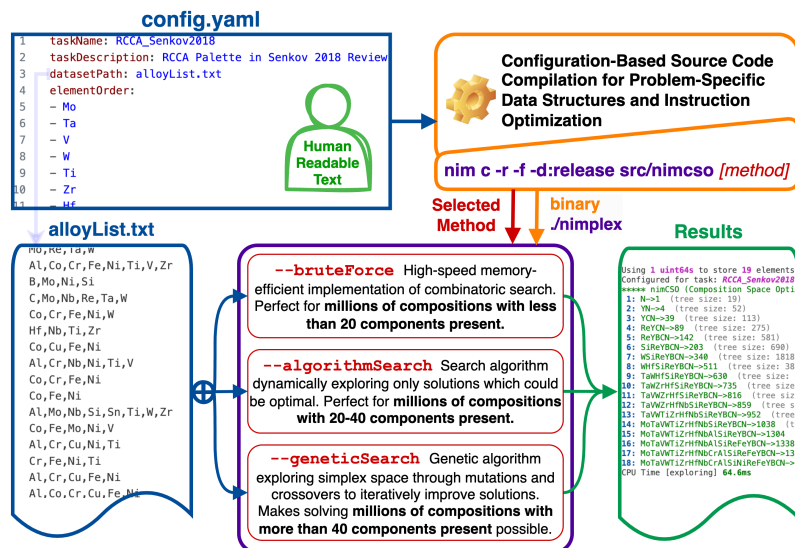
Methods and Performance

Overview

Figure 1: Schematic of core nimCSO data flow with a description of key methods. Metaprogramming is used to recompile the software optimized to the human-readable data and configuration files at hand.

Acknowledgements

References

Cantor, B., Chang, I. T. H., Knight, P., & Vincent, A. J. B. (2004). Microstructural development in equiatomic multicomponent alloys. *Materials Science and Engineering A*, *375-377*, 213–218. https://doi.org/10.1016/j.msea.2003.10.257

Debnath, A., Krajewski, A. M., Sun, H., Lin, S., Ahn, M., Li, W., Priya, S., Singh, J., Shang, S., Beese, A. M., Liu, Z.-K., & Reinhart, W. F. (2021). *Journal of Materials Informatics*, *1*. https://doi.org/10.20517/jmi.2021.05

Krajewski, A. M., Beese, A. M., Reinhart, W. F., & Liu, Z.-K. (2024). *Efficient generation of grids and traversal graphs in compositional spaces towards exploration and path planning exemplified in materials*. http://arxiv.org/abs/2402.03528

Li, W., Ahn, M., Raman, L., Debnath, A., Lin, S., Krajewski, A. M., Shang, S., Priya, S., Reinhart, W. F., Liu, Z.-K., & Beese, A. M. (2023). Design and validation of refractory alloys using machine learning, CALPHAD, and experiments. *(Under Review in International Journal of Refractory Metals and Hard Materials)*. https://doi.org/10.2139/ssrn.4501281

Yeh, J. W., Chen, S. K., Lin, S. J., Gan, J. Y., Chin, T. S., Shun, T. T., Tsau, C. H., & Chang, S. Y. (2004). Nanostructured high-entropy alloys with multiple principal elements:

Novel alloy design concepts and outcomes. *Advanced Engineering Materials*, *6*, 299–303. https://doi.org/10.1002/adem.200300567