An Arc Density Maximum Flow Algorithm

Ron Shepard[1], Scott R. Brozell[1], Jeffrey Larson[2], Paul Hovland[2], Sven Leyffer[2]

[1]Chemical Sciences and Engineering Division, Argonne National Laboratory, Lemont, IL 60439
[2]Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439

The arc density is a compact graphical depiction of a molecular electronic structure wave function based on the Shavitt graph within the graphical unitary group approach (GUGA)[1]. The Shavitt graph is a hierarchical directed acyclic graph (DAG), whose nodes k depend on the quantum numbers n_k (the orbital level), N_k (the number of electrons), and S_k (the $\hat{S}^2$ quantum number). The arcs connect the nodes in adjacent levels within this graph. Each configuration state function (CSF) corresponds to a path from the tail to the head of the graph, touching a single arc between each level. A wave function may be expanded in the basis of these CSFs, and the arc density of a particular arc is the sum of squares of the coefficients of all CSFs that touch that arc.

An algorithm for computing the maximum-flow path in a network is applied to the identification of the dominant CSF in a given wave function. The recursive algorithm loops over the nodes of the arc density at each level, and, for each node, loops over its incoming arcs, of which there are at most four, determining an upper bound to the maximum possible incoming flow for that node. The maximum squared CSF coefficient is identified with the maximum possible incoming flow. The walk corresponding to that CSF is generated by backtracking from the graph head to the graph tail. Elimination of the arc density contributions for that CSF from its arc path, and reapplication of the algorithm, produces a sequence of CSFs in decreasing order of coefficient magnitudes.

For relatively small CSF expansions, this ordering can be determined simply by sorting/ranking the CSF coefficient vector, requiring at most $O(N_{\mathrm{CSF}} \log(N_{\mathrm{CSF}}))$ effort depending on the number of CSFs sought. However, for large CSF expansions, this becomes impractical (e.g. Ref. 2 includes an expansion with $N_{\mathrm{CSF}} \approx 10^{150}$). The maximum-flow algorithm requires $O(MN_{\mathrm{node}})$ effort where M is the number of coefficients sought and N_{node} is the number of nodes in the Shavitt graph. For a full-CI wave function expansion, $N_{\mathrm{node}} \approx O(N_e^2 n)$ where N_e is the number of electrons and n is the total number of orbitals; other expansion types have even fewer nodes[2], e.g. $O(N_e n)$ or even $O(n)$. Due to this scaling, the maximum-flow algorithm is practically applicable to wave function expansions of any size.

[1] R. Shepard, S. R. Brozell, and G. Gidofalvi, *J. Computational Chem.* **41**, 129-135 (2020). DOI: 10.1002/jcc.26080.

[2] R. Shepard and S. R. Brozell, *Mol. Phys.* **117**, 2374-2390 (2019). DOI: 10.1080/00268976.2019.1635275.

This work was performed under the auspices of the Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences, U.S. Department of Energy, under contract number DE-AC02-06CH11357.