

# Chem 279 Final Project

Philo Beshay | James McTighe

# Project Goals Outlined

1. Optimize molecular geometries using steepest descent approach.
2. Determine vibrational frequencies of molecules using CNDO/2.
3. Using calculated results and generate infrared spectra utilizing pybind.
4. Investigate the limitations of the CNDO/2 method and basis sets.

# Theory

Vibrational Frequencies of a molecule can be calculated via CNDO/2

- Second Derivative with respect to position (Force Constants)
- Mass weighted hessian matrix
- Eigen values correspond to Frequencies
- Eigen vectors are vibrational normal coordinates

Hessian Matrix:

$$\frac{\partial E^2}{\partial^2 R} = F$$

Mass Weight:

$$G_{ij} = \frac{1}{\sqrt{m_i m_j}}$$

Main Function:

$$\left[ - \sum_{i=1}^{3N_{\text{atom}}} \left( \frac{1}{2} \frac{\partial^2}{\partial q_i^2} \right) + \frac{1}{2} \mathbf{q}^t (\mathbf{U} (\mathbf{F} \cdot \mathbf{G}) \mathbf{U}^t) \mathbf{q} \right] \Psi_{\text{nuc}} = E_{\text{nuc}} \Psi_{\text{nuc}}$$

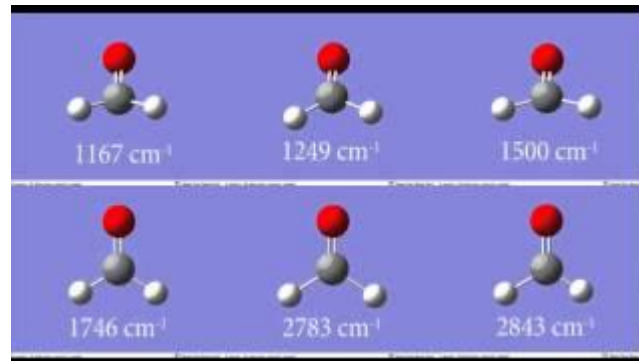
Eigenvalues:

$$\nu_i = \frac{1}{2\pi} \sqrt{\epsilon_i}$$

*q => eigenvectors*

# Key Considerations

- First Derivative is zero (optimized geometries)
- Eigenvalues include translational and rotational frequencies
  - Translations  $\approx 0.01 \text{ cm}^{-1}$
  - Rotation  $\approx 10\text{-}50 \text{ cm}^{-1}$
  - Rotation is higher because geometry can't be optimized to give an exact gradient of zero. Cannot be more accurate than  $0.1 \text{ cm}^{-1}$  (1)
- Typically, translational and rotational frequencies are removed via projection. However, we did not perform that here.
- If these extra movements are of a significant magnitude, they will “mix” with the true vibrational results and yield inaccurate results.



# Geometry Optimization

## Approach

We use steepest descent algorithm with line search, the algorithm runs iteratively and each iteration:

1. Computes total nuclear + electronic energy via SCF.
2. Computes gradient of energy.
3. Moves atoms in the direction negative to gradient.
4. Uses golden-line search to find best step size.
5. Stops once gradient is below designated threshold (or once max iterations reached)

# Geometry Optimization

## Implementation

Computation done in C++ using pybind wrapper to create class module for visualization.

Class is initialized from xyz atom file and includes features such as:

- `Get_geometry()`: prints geometry of molecule as list of tuples in format (Z, x, y, z). Can be used before & after optimization.
- `Get_energy()`: print total energy in eV. Can also be used before & after optimization for comparison.
- `Optimize()`: Runs steepest descent geometry optimization.

# Geometry Optimization

## Results

Molecule	Energy before optimization	Energy after optimization	Difference	Num iterations	Time to optimize
H2	- 30.7765 eV	- 40.57527 eV	- 9.79877	100	0.6s
H2O	-169.255a1 eV	- 505.475 eV	- 336.2199 (Not converged)	50	28m15s

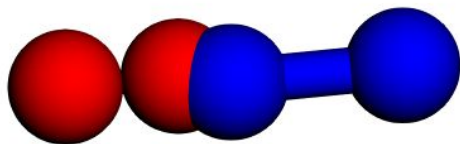
# Geometry Optimization

## Results - Visualization (Py3Dmol)

H<sub>2</sub> geometry comparison:

Red - Optimized with steepest descent

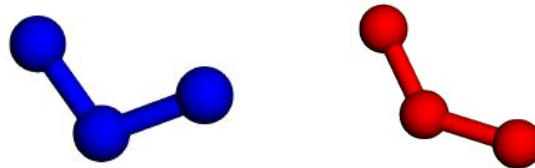
Blue - Geometry from Materials Project



H<sub>2</sub>O geometry comparison:

Red - Optimized with steepest descent

Blue - Geometry from Materials Project





# Geometry Optimization

## Limitations & Conclusion

Approach is not successful due to various reasons:

1. Very computationally expensive, running full SCF energy + gradient calculation at each iteration.
2. Modern semi-empirical methods (e.g., MP2, DFT) much better for geometry calculations.
  - a. Due to not accounting for electron correlations, only valence electrons are considered.
  - b. STO-3G basis set is very simplified and often leads to geometry errors specifically, larger basis sets are often used for these calculations (e.g., 6-31G\*, cc-pVDZ). -> approach usually underestimates bond length. (2, 4, 5)
3. Model implementation does not accurately converge, especially for larger molecules.

# Geometry Optimization

## Next Steps

1. Complete optimization for other molecules. (OH, HF, possibly larger structures) & visualized geometry comparisons. (Before vs After vs MP Geometry)
2. Look for workaround to reduce the computational complexity of the task for larger molecules, such as using two tier-gradient threshold, comparing against the density matrix & only calculating full SCF per iteration once the first gradient threshold is reached.

Due to suboptimal results of this method & the computational cost, we decided to calculate IR spectra using the optimized geometries from Materials Project for these molecules.

# Main Source Functions

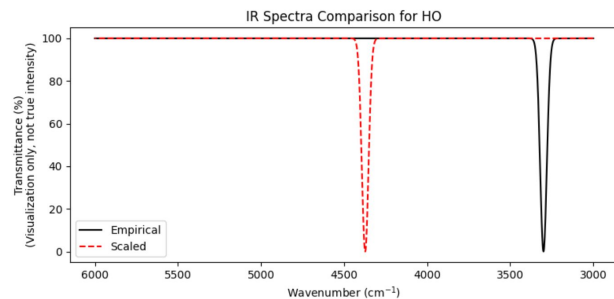
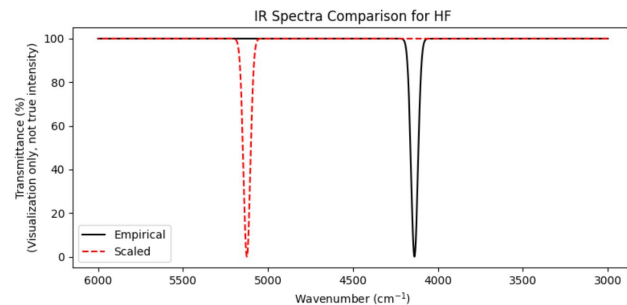
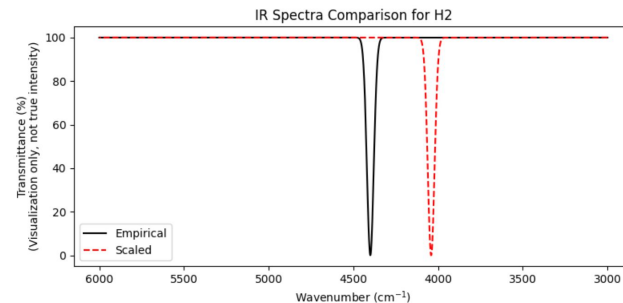
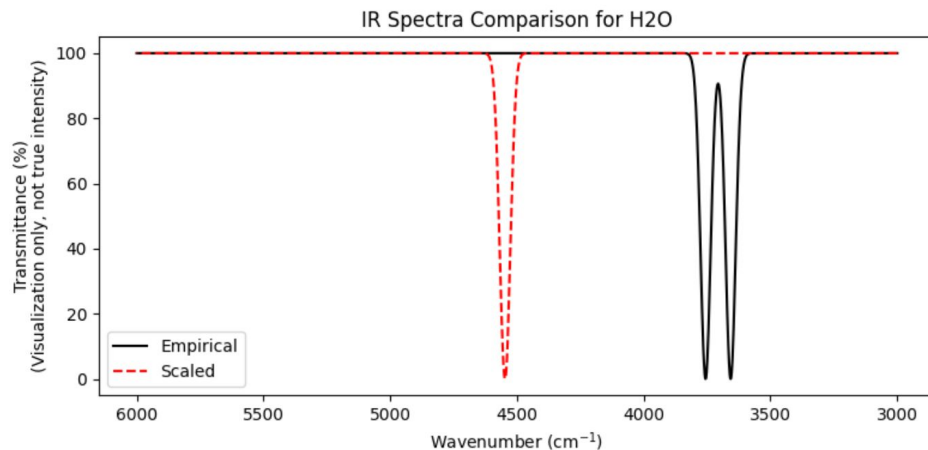
- The hessian matrix is built utilizing a central difference function which iteratively evaluates the gradient after making alterations for each coordinate of each atom.
  - Default step size of  $h = 1.0 \times 10^{-3}$
- Hessian Matrix is weighted by mass using G matrix
  - Per Shaw et al (2) we scaled the force constant by a factor of 1.48
- Eigenvalues (vibrational frequencies) and Eigenvectors (vibrational normal coordinates) are calculated using armadillo's eig\_sym function.
- The largest Eigenvalues correspond to the expansion / contraction.
  - Other values are the rotational and translational movements

# Data

Frequencies and Percentage Difference From Empirical values				
	H <sub>2</sub>	H <sub>2</sub> O	HF	HO
Calculated cm <sup>-1</sup>	4917	5515, 5524	6234	5320
Calculated (Scaled) cm <sup>-1</sup>	4041	4553, 4541	5124	4373
Empirical cm <sup>-1</sup>	4400	3657, 3756	4138	3300
Percent Difference (un-Scaled)	11%	50%, 47%	51%	61%
Percent Difference (Scaled)	8%	25%, 21%	24%	33%

# Generated Graph Examples

Overlap of 2 peaks



# Discussion - Basis Sets

- In general the results we obtained were poor
  - Largely limited by the quality of our model
- After performing some literature review, we determined that STO-3G is an inadequate basis set for modeling vibrational frequencies as compared with other basis sets, such as 3-21G, as per Shaw et al.(2)
  - "...the STO-3G frequencies and particularly the vibrational eigenvectors are completely unreliable."
  - Scaling using least-squares optimization helps to improve results
  - A 3-21G basis set is more flexible because it contains twice as many valence functions that can combine freely to model Molecular orbitals (1)
  - An optimal basis set for this problem would be one that is *Polarization Consistent* to account for shifting electron density as the bonds expand and contract.
- Shaw et al, complimenting our results, notes that STO-3G basis sets significantly overestimate frequencies
- Other sources, such as Schlegel et al (3) agree that STO-3G offer poor performance for computing force constants and suggest 4-31G as a substitute.

# Discussion - Geometry Optimization

- An additional source of error in our current calculations stems from the presence of notably high rotational frequencies ( $\sim 500 \text{ cm}^{-1}$ ).
- These elevated frequencies manifest prominently within the computed eigenvalues, serving as a clear indicator of un-optimized molecular geometries. This sub-optimal structural representation introduces inaccuracies that propagate through the subsequent analysis and results.

# Conclusions

We extended our existing code used in previous problem sets to solve an additional application of Computational Chemistry. This involved adapting the existing framework to handle new molecular systems or a different type of calculation (e.g., optimization, frequency analysis, or property prediction) that was not the focus of earlier assignments.

While our results were not desirable in terms of their accuracy, this allowed us an opportunity to expand our knowledge on the theory behind these methods. Specifically, the discrepancies between the calculated and expected/experimental values led to a deeper investigation of the underlying approximations, the limitations of the chosen basis set, and the level of theory employed. Analyzing these errors was a crucial learning experience in understanding the trade-offs and practical considerations in computational modeling.



# References

- (1) Jensen, F. *Introduction to Computational Chemistry*; New Jersey John Wiley & Sons Inc, 2017.
- (2) Shaw, R. A.; Ursenbach, C.; Rauk, A.; Wieser, H. Comparison of STO-3G and 3-21G *Ab Initio* Harmonic Force Fields for Ethane, Propane, Dimethyl Ether, and Cyclobutane: Effects of Geometry and Scaling on Calculated Frequencies, Eigenvectors, and Infrared Absorption Intensities. *Canadian Journal of Chemistry* **1988**, 66 (5), 1318–1332. <https://doi.org/10.1139/v88-214>.
- (3) Bernardi, F.; Schlegel, H. Bernhard.; Wolfe, S. Ab Initio Computation of Force Constants. *Journal of Molecular Structure* **1976**, 35 (1), 149–153. [https://doi.org/10.1016/0022-2860\(76\)80110-x](https://doi.org/10.1016/0022-2860(76)80110-x).
- (4) Bálint, D.; Jäntschi, L. *Comparison of Molecular Geometry Optimization Methods Based on Molecular Descriptors. Mathematics* **2021**, 9 (22), 2855. [Comparison of Molecular Geometry Optimization Methods Based on Molecular Descriptors](#)
- (5) Gulania, S.; Whitfield, J. D. *Limitations of Hartree–Fock with Quantum Resources. Journal of Chemical Physics* **2021**, 154 (4), 044112. [Limitations of Hartree–Fock with quantum resources](#)