MMMGRUBS: MOLECULAR MOTION MOVIES AND GEOMETRY RECONSTRUCTION USING BAYESIAN STATISTICS

by

ALI RAMADHAN

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Physics

Waterloo, Ontario, Canada, 2017

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

MMMGRUBS: MOLECULAR MOTION MOVIES AND GEOMETRY RECONSTRUCTION USING BAYESIAN STATISTICS

ALI RAMADHAN

Hakuna matata?

April 5, 2017

ABSTRACT

Ever since the early days of Coulomb explosion imaging, the direct imaging of the structure and dynamics of individual molecules has been promised; however, in practice no one has been able to accurately retrieve this structure and all imaging relies on plotting the momentum vectors of the fragments in different ways to infer crude changes in geometry. The momentum vectors tell a large part of the story but do not provide a clear picture of the molecular dynamics everyone seeks so an accurate method of retrieving the structure is highly desirable. I use Bayesian inference and Markov chain Monte Carlo methods to elucidate the molecular geometries, allowing for the study of larger systems than before and for the inclusion of measurement error in these studies for the first time. Using this method, the ultrafast isomerization of carbonyl sulfide and acetylene is imaged in position-space and showcased as a molecular movie. The method, however, is much larger in scope and is generalized to any problem involving a physical system described by a system of ordinary differential equations, where the final conditions are experimentally measured and inference of the initial conditions is desirable.

ACKNOWLEDGMENTS

Hi everyone, thanks!

Look mom and dad! I got my name on a thesis!

Also dedicated to you, yes you the reader, for reading this thing!

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

1

INTRODUCTION

To image the dynamics of a molecule by destroying it seems paradoxical. As we shall see, the molecule's structure is encoded in the atmoic shrapnel left behind after an explosion. However, to destroy one of nature's simplest creations is no easy task. Molecules are held together by strong chemical bonds. Our best line of attach is to shoot them with a short laser pulse—engulfing the molecule in an intense oscillating electric field will strip away its electrons and cause it to break up quickly. In these first few chapters, I will discuss how to create a short laser pulse and how these pulses interact with matter. Then I will introduce the technique of pump-probe Coulomb explosion imaging, which we will use to probe the dynamic structure of small molecules by studying the atomic fragments resulting from the explosion.

2

A BRIEF HISTORY

2.1 HISTORY AND ACCOMPLISHMENTS

The idea that Coulomb explosion imaging could be used to produce so-called molecular movies has permeated the literature ever since the technique's emergence. The goal of observing

2.1.1 *Coulomb explosion imaging*

CEI was first performed by Vager, Naaman, and Kanter (1989) whereby the Coulomb explosion was initiated by passing the molecule through a thin carbon film at high velocities.

There exist other methods of initiating CEI, among them thin foils, free-electron lasers, highly-charged ion impact, single X-ray photons from a synchrotron source.

2.1.1.1 *Molecular movies in momentum-space*

2.1.1.2 *Geometry reconstruction*

Surprisingly, an attempt was made to arrive at an analytical solution for calculating geometries from measured momentum data. Nagaya and Bandrauk (2004) are able to derive so-called classical imaging formulas giving the position wavefunction squared for the Coulomb explosion of a diatomic molecule and a linear triatomic molecule (the cases of symmetric and asymmetric Coulomb explosion are treated).

They seem to have worked hard to find an analytical solution but their unsaid conclusion seems to be that it is an intractable problem and their group went silent on this problem.

Légaré et al. (2005a,b) was the first to use Coulomb explosion imaging and report molecular structures. To obtain structures, they assume the explosion proceeds under a purely Coulombic potential and use optimization methods to make guesses at the structure that most accurately reproduces the observed data consistent with minimizing a least-squares objective function. Unfortunately they provide very minimal information regarding their methods and there is a complete lack of discussion acknowledging the shortcomings of this method. Using 8 fs laser pulses they report on the structure of D2O and SO2 (Légaré et al., 2005b). They also claim to have imaged vibrating D2+ and dissociating SO22+ and SO23+ however they provide no more than a couple of dissociation frames and infer the transient D2+ bond length from kinetic energy release ratios as a function of pump-probe time delay (Légaré et al., 2005a).

The main shortcomings being degenerate solutions and the fact that they employ convex optimization methods to a problem that is not convex. It is not clear if they even knew about these issues.

Gagnon et al. (2008) reported the reconstruction of dichloromethane (CH2Cl2) using a home-made stochastic-based simulated annealing

There is nothing wrong with writing your own code here but nonconvex optimization

algorithm that globally optimizes the molecular spatial configuration. They discuss uncertainties but are only able to obtain the structure in five cases. [1]

The best effort so far has probably been the one by KUNITSKI et al. (2015) in which they use a lookup table approach to image the Efimov state of the helium trimer.

2.1.2 *Laser-induced electron diffraction*

2.1.3 *Femtosecond X-ray scattering*

2.1.4 *In the biological sciences*

Molecular movies are of course not only of interest in physics and chemistry as a means of probing fundamental processes, but also in the biological sciences where molecular structure play a crucial role in determining the function of biomolecules such as proteins. However, the molecules of interest there are much too large to be studied by any of the previous techniques. Thus molecular movies in the biological sciences tend to be annotated computer simulations amalgamated from multiple studies. That said, they are very impressive pieces of work.

A particularly impressive movie by CHEUNG and CRAMER (2012) showcases the process of RNA polymerase transcription and goes on for over six minutes.

1 See BOCHAROVA et al. (2011) for a discussion of optimization algorithms.

3

COULOMB EXPLOSION IMAGING

3.1 EXPERIMENT OUTLINE

3.1.1 *Pump-probe Coulomb explosion imaging*

In pump-probe Coulomb explosion imaging (CEI) one ultrashort laser pulse is split into two pulses through the use an asymmetric beam-splitter. One of the pulses, the pump pulse, is usually much weaker than the other, the probe pulse. A time delay τ between the pulses is then created such that the pump pulse goes first and the probe pulse second. The job of the pump pulse will be to initiate some change in the molecule. One example could include an isomerization of the molecule. Thus the pump pulse "pumps" the molecule into some excited state. The job of the powerful probe pulse is to engulf the molecule in an intense enough laser field such that multiple electrons are stripped off of it. The molecule's individual atoms are left in a highly-charged state and begin to behave as individual point charges in a purely Coulombic potential. The entire process occurs in the presence of a constant electric field and so the positively-charged ions all accelerate upwards towards a time- and position-sensitive detector. Thus the probe pulse allows for the "probing" of the excited state.

3.1.2 *Femtosecond Multiple Pulse Length Spectroscopy*

3.2 EXPERIMENTAL APPARATUS

3.3 TIME AND POSITION MEASUREMENT

The position is then calculated using

$$x = \frac{Q_1 + Q_2}{Q_1 + Q_2 + Q_3 + Q_4}, \quad y = \frac{Q_1 + Q_3}{Q_1 + Q_2 + Q_3 + Q_4} \tag{3.1}$$

3.4 CALCULATING ATOMIC FRAGMENT MOMENTA

The components of the three-dimensional momentum vector $\mathbf{p} = (p_x, p_y, p_z)$ for each atom are then calculated as

$$p_x = \frac{m(x - x_0)}{t}, \; p_y = \frac{m(y - y_0)}{t}, \; p_z = \frac{qV}{2\ell}\left(\frac{t_0^2 - t^2}{t}\right) \tag{3.2}$$

3.5 UNCERTAINTY IN MOMENTUM MEASUREMENTS

For any relation $f = f(x_1, x_2, \ldots, x_n)$, assuming independent variables, the absolute uncertainty in f is

$$df = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} dx_i \right)^2} \tag{3.3}$$

3.6 SIMULATING THE COULOMB EXPLOSION

To simulate an explosion of a molecule containing n atoms, we must solve the classical equations of motion for each ion right after the explosion. We choose to use Hamiltonian mechanics here to acquire a system of first-order differential equations which may be easily solved by numerical methods such as the ubiquitous fourth-order Runge-Kutta. Assuming a purely electromagnetic potential for each ion, the Hamiltonian of the molecular system is

$$\mathcal{H}(\mathbf{r}_i, \mathbf{p}_i, t) = \sum_{i=1}^{n} \frac{\mathbf{p}_i^2}{2m_i} + \frac{1}{4\pi\epsilon_0} \sum_{\substack{\{i,j\} \\ i \neq j}} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}$$

where the second summation is over all i, j pairs where $i \neq j$.

Calculating Hamilton's equations for the system, we get

$$\dot{\mathbf{r}}_i = \frac{\partial \mathcal{H}}{\partial \mathbf{p}_i} = \frac{\mathbf{p}_i}{m_i}$$

$$\dot{\mathbf{p}}_i = \frac{\partial \mathcal{H}}{\partial \mathbf{r}_i} = \frac{1}{4\pi\epsilon_0} \sum_{i \neq j} \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3}$$

where i is held fixed over the second summation.

4

GEOMETRY RECONSTRUCTION USING A LOOKUP TABLE

Having destroyed a molecule and measured the momentum of each of its atomic fragments, we are left with the inverse problem of inferring its structure. While explosions proceed in a deterministic fashion, that is structures map bijectively to momentum measurements, the converse is not true. Two very different structures may produce the same momentum measurements. To make matters worse, there is no analytic solution to the problem and as the molecule grows, the problem of finding its structure becomes increasingly high dimensional. To combat this problem we will require the use of various mathematical and statistical methods. I first discuss some results from the theory of inverse problems to shed some general insight on these problems. I then follow with a discussion of optimization methods which may be used to tackle the problem for very small molecules. However, for full imaging of larger polyatomic molecules with an analysis of measurement error, Bayesian inference using Markov chain Monte Carlo methods is the way to go, which I discuss in the last chapter of this part.

4.1 LOOKUP TABLE

In this approach, many Coulomb explosions are simulated for a wide variety of structures, and the resulting momentum vectors from each simulation are stored. Thus you have a mapping from molecular structures to momentum vectors. To determine the structure belonging to a certain set of observed momentum vectors, you simply read the table in reverse. This approach is simple to implement, very quick by design, and front-loads the computation which may be desirable for large data sets. However, of course, it has an exponential time and space complexity $\mathcal{O}(e^{3N-6})$ where N is the number of atoms.

4.2 RESULTS FOR OCS

5

GEOMETRY RECONSTRUCTION AS AN OPTIMIZATION PROBLEM

Much progress was made over the lookup table by treating the geometry reconstruction problem as a constrained nonlinear convex optimization such that MATLAB's `fmincon` function an be relied on. It relies on trust regions and uses an interior-point algorithm. This worked especially well in the case of triatomic molecules however four-atom systems proved incredibly difficult to tackle here. This was due to the exponential increase in the number of saddle points with dimensionality making the problem highly non-convex and unsuitable for `fmincon`.

Recall that triatomic molecules have three degrees of freedom resulting in a problem of dimension 3 while four-atom systems have six. That is, $3N + 6$ for an N-atom system.

5.1 SIMPLEX ALGORITHM

Brichta, Seaman, and Sanderson (2009) proposed the reconstruction of small triatomic molecules using a simplex algorithm. Unfortunately they only report on the reconstruction of molecular structures based on simulated data for carbon dioxide and formaldehyde. I could not use this algorithm to find the geometries of CO_2 or OCS from real data.

5.2 RESULTS FOR OCS

BIBLIOGRAPHY

Bocharova, I. et al. (2011). Charge Resonance Enhanced Ionization of CO_{2} Probed by Laser Coulomb Explosion Imaging. *Physical Review Letters* **107**, 063201–063201. DOI: 10.1103/PhysRevLett.107.063201.

Brichta, J.-P., Seaman, A. N., and Sanderson, J. H. (2009). Ultrafast imaging of polyatomic molecules with simplex algorithm. *Computer Physics Communications* **180**, 197–200. DOI: 10.1016/j.cpc.2008.09.006.

Cheung, A. and Cramer, P. (2012). A Movie of RNA Polymerase II Transcription. en. *Cell* **149**, 1431–1437. DOI: 10.1016/j.cell.2012.06.006. (Visited on 06/06/2016).

Gagnon, J, Lee, K. F., Rayner, D. M., Corkum, P. B., and Bhardwaj, V. R. (2008). Coincidence imaging of polyatomic molecules via laser-induced Coulomb explosion. *Journal of Physics B: Atomic, Molecular and Optical Physics* **41**, 215104. DOI: 10.1088/0953-4075/41/21/215104. (Visited on 09/26/2016).

Kunitski, M. et al. (2015). Observation of the Efimov state of the helium trimer. en. *Science* **348**, 551–555. DOI: 10.1126/science.aaa5601. (Visited on 06/28/2016).

Légaré, F., Lee, K., Litvinyuk, I., Dooley, P., Bandrauk, a., Villeneuve, D., and Corkum, P. (2005a). Imaging the time-dependent structure of a molecule as it undergoes dynamics. *Physical Review A* **72**, 052717. DOI: 10.1103/PhysRevA.72.052717.

Légaré, F. et al. (2005b). Laser Coulomb-explosion imaging of small molecules. *Physical Review A* **71**, 013415. DOI: 10.1103/PhysRevA.71.013415.

Nagaya, K. and Bandrauk, A. D. (2004). Laser Coulomb explosion imaging of linear triatomic molecules. *Journal of Physics B: Atomic, Molecular and Optical Physics* **37**, 2829–2853. DOI: 10.1088/0953-4075/37/14/001. (Visited on 08/29/2016).

Vager, Z., Naaman, R., and Kanter, E. P. (1989). Coulomb explosion imaging of small molecules. *Science* **244**, 426–431. This is a classic of linguistics.

INDEX