Systematic Variation of Molecular Properties Expressed Through Quantum Control Landscapes

Richard Li

September 21, 2009

1 Introduction

Over the course of the summer I worked in the general area of expressing molecular and material properties as optimal control problems. Although it may seem unusual to express molecular or material property optimization as a control problem, in fact, it is in fact quite natural and beneficial, as shown in this research. If a property, J, is treated as an observable of a quantum system, it is then possible to optimize J in terms of appropriate combinations of the "controls". In this work, the controls are simply different combinations of moieties on a molecular scaffold. The relationship between J and the controls defines an *optimal control landscape*. Due to the inherent nature of quantum systems, the topology of the landscape should ideally be *trap free*, meaning there should be no sub-optimal extrema [1]. The present research used various forms of molecular spectroscopy to validate this rather surprising fact. Molecular spectroscopy is not only a powerful technique that is often used for structure elucidation, but, more importantly for this research, spectra provides an observable property based on complex electronic and steric interactions in a molecule. As will be shown below, the landscapes validated these expectations concerning a trap-free topology. The existence of a trap-free control landscape lends itself to special predictive power within the space of the controls and can also be used as an aid in the optimization of molecular based material properties.

2 Background

The unintuitive fact that a trap-free control landscape exists is a consequence of the inherent nature of quantum systems under control [1]. To characterize an m-level open quantum system of an ensemble of molecules that interacts with the environment, it is convenient to introduce the density matrix ρ. The density matrix ρ describes the statistical state of the system; i.e., the probability that the molecules will be in any of their m levels. These matrices must necessarily have unit trace, and, they are also positive. As a result, the set D of all density matrices ρ, is a convex set [2]. The objective function J, which specifies the observable, is defined by $J = Tr(\rho O)$ [3], where O is an appropriate Hermitian quantum observable operator. Because J is linear in ρ, it is a convex function (or equivalently, -J a concave function). It is well-known that convex functions have only one minimum (or a maximum, if concave). Thus, the landscape is trap-free, with no sub-optimal points, since there is only one global extremum value. Note that it is not necessary to know the explicit form of ρ; as long free access is available to a complete set, D, then a trap-free landscape will exist. Thus, a trap-free landscape is guaranteed, provided the accessible set of density matrices is convex (the controls are sufficiently unconstrained) and the choice of a reasonable target objective [1].

3 Methods

The use of spectral data as a property for testing the theorized trap-free moleculary property landscape was facilitated by the ready availability of a series of spectral databases, Know-It-All-U, distributed by Bio-Rad, that provided access to hundreds of thousands of spectra. The work performed this summer was focused on using ^{1}H and ^{13}C NMR (nuclear magnetic resonance) data to create landscapes for molecular scaffolds with two variable sites. For example, consider a para-substituted benzene (see Figure 1), where the moieties bonded at sites X and Y form the "controls", as explained below.

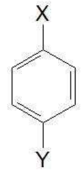

Figure 1: Para-substituted Benzene Ring

NMR shifts of a particular ^{1}H or ^{13}C atom extracted from the spectra primarily reflect the shielding effect of the surrounding electrons; i.e., the shift depends on the electron density around the particular atom. The greater the electron density, the greater the shielding effect, and the lower the recorded shift. The spectroscopic shift (J value) was collected from the database as a function of the moieties on each of the two sites, X and Y. Consider a landscape created using ^{1}H NMR data for a para-substituted benzene. For ^{1}H NMR on a para-substituted benzene, there is only one unique position due to symmetry - the hydrogen adjacent to a moiety (either X or Y). Any hydrogen on the benzene ring sees a moiety on the adjacent carbon (in the "ortho position"), or two carbons away (in the "meta" position). Given n_1, and n_2 different moieties for site X and site Y, respectively, the J value was collected for every available combination out of the possible $n_1 \times n_2$ cases (databases understandably do not often contain every single desired molecule in the full set). Each of the n_1 moieties found on the X position is randomly assigned an integer label: the first moiety is assigned 1, the second 2, and so on. The same is done for the moieties on the Y position. In this way, the spectral property as a function of the discrete variables on the associated x and y axes can be constructed. The observed shift is then plotted as a function of these two variables. [4][5] (see Figure 2).

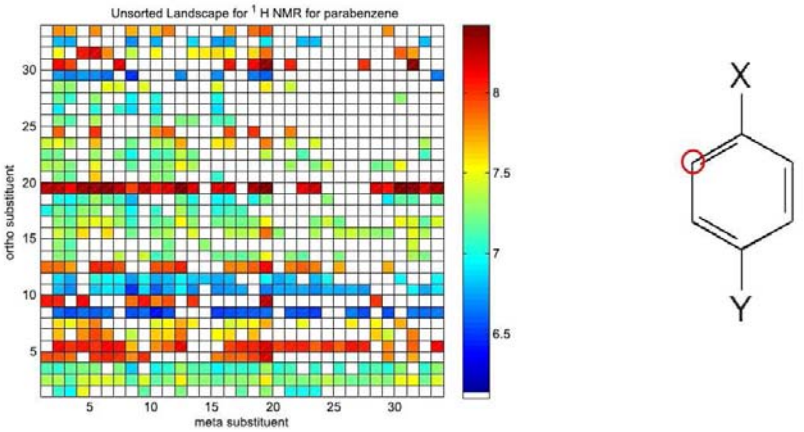

Figure 2: Unsorted ^{1}H NMR data for para-substituted benzene taken at the red circle. The x and y axes correspond to the random labeling of the moieties on the two sites. "Ortho" corresponds to the X position, while "meta" corresponds to the Y position.

The landscape depicted in Figure 2 apparently contradicts the theoretical prediction concerning a trap-free character. However, this contradiction is quickly resolved through the introduction of a reordering algorithm to find an optimal labeling [5], one that shows the trap-free nature of the landscape (shown later). As noted, each moiety was assigned a unique integer. However, the choice of integer labeling was random. If the database was ordered differently, a different labeling of the x-axis and y-axis would be obtained. Therefore, a reordering algorithm is necessary to produce the "correct" ordering that reveals the true regularity of the landscape. An ordering that gives a trap-free landscape is expected because of the convexity of D, the set of all density matrices, ρ. There should be a systematic ordering based on physical grounds, and, if the appropriate physical rules are known, it would be possible to construct an optimal ordering based on knowledge of the moieties. For example, there are several well-known general "rules" for NMR founded on chemical properties concerning the distribution of electron density (e.g., oxygen is a "donor" group that denotes its electrons, thus causing greater electron shielding). Knowledge of such rules could be used to manually create an order of moieties that would generate a regular landscape. However, the reordering algorithm employed in this work does not require any such prior knowledge and is still able to produce an ordering that ultimately yields a trap-free landscape, thereby confirming the original hypothesis. For the work performed this summer, a very simple sorting algorithm was used, although more advanced ones exist for treating more complex cases[5]. Here, we define Q_m as a measure of the monotonocity of the landscape for the mth moiety on site X,

$$Q_m = \frac{1}{N} \sum_{m' \neq m} \sum_{n} \left(a_{mn} - a_{m'n} \right), \qquad \forall a_{mn}, a_{m'n} \neq 0, \tag{1}$$

where, n is an index of site Y, m is the index of X, a_{mn} and $a_{m'n}$ are the measured shift values for moiety m or m$'$ on site X, and n on site Y, and N is the number of molecules compared along the Y position for all $m' \neq m$ (we ignore comparisons if either a_{mn} or $a_{m'n}$ is equal to 0). The optimal ordering of the moieties on site X is then determined by sorting the n_1 values of Q_m, independent of site Y. The same is then done for site Y. The ordering produced by this algorithm fully coincided with well known rules of NMR. Figure 3 shows an example of the reordering algorithm in operation.

Using this algorithm to find a suitable order of the moieties with the data in Figure 2 produced a regular landscape in Figure 4. This method can be applied to many scaffolds, such as benzenes, alkenes,

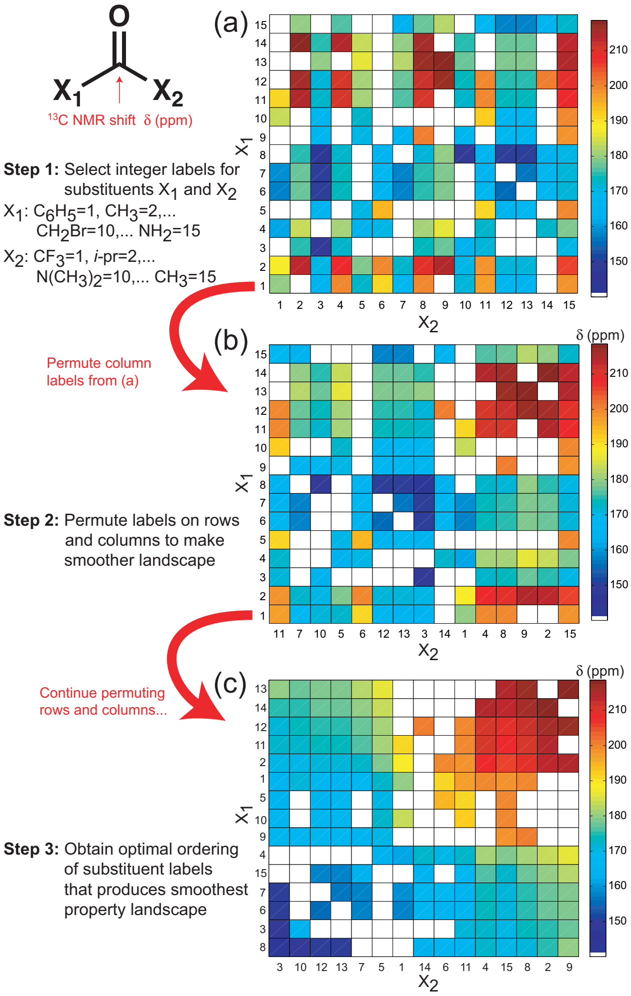

Figure 3: Operation of the reordering algorithm for a carbonyl scaffold with ^{13}C NMR shift as the property. (a) shows the initial (random) order chosen for the moieties. Some columns are permuted in (b) to show that suitable permutation makes the landscape appear ”better” (more regular). Permuting the remaining columns and rows gives (c). The white spaces in the landscape represent combinations of moieties for which no NMR data was available.

carbonyls, and other simple molecules with each producing its own landscape. Note that although the example above had only two variable sites, a trap-free landscape is expected to arise for applications with an an arbitrary number of sites (variables).

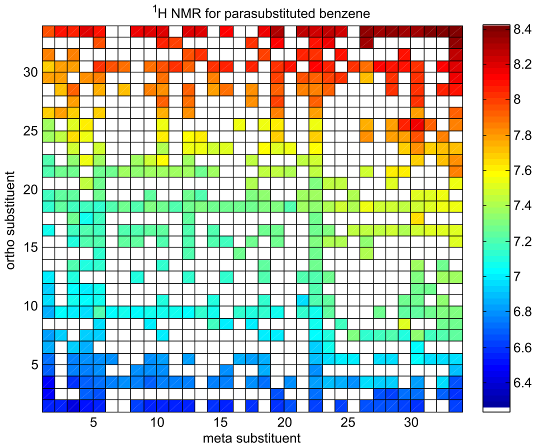

Figure 4: Reordered ^{1}H NMR data for para-substituted benzene. This regular reordered landscape should be compared to its original random parent in Figure 2.

4 Findings

The fifteen landscapes that were generated on various scaffolds show quite definitively that NMR as a spectroscopic property is fully consistent with a control theory perspective. Also, the optimal ordering that was obtained often corresponded to well-known rules of NMR. We were thus able to generate results consistent with accepted chemical knowledge without imposing any prior bias towards such chemical regularity. Furthermore, these landscapes also have a clear predictive capability, behavior that Bio-Rad was very interested in. Within the search space of the landscape generated, it is easy to do a local interpolation to predict the gaps (the white space) in the landscapes. Therefore, given an incomplete, or even somewhat sparse, set of data, it is possible to create a reasonable landscape that can be interpolated for unknown spectral data. Such a predictive capability can be valuable for analytical purposes. Preliminary tests show reasonable prediction quality. As an example, Figure 5 shows a truth plot of predicted values versus actual values for ^{13}C NMR data for a para-substituted benzene (Figure 1). The standard deviation of the predicted values from the actual values is 1-3ppm. Figures 6 and 7 show other NMR shift landscapes. For all fifteen cases studied, similar predictive capabilites were found. Any molecular property, besides spectra, can be viewed from a control perspective with the associated landscapes generated in the same manner. Those landscapes can then be used as a guide in the creation of molecules to satisfy a particular molecular based material property.

5 Conclusion

This work centered on ^{1}H and ^{13}C NMR shifts which generally follow well-known rules. Future work will apply similar techniques to more complex spectroscopic data, such as IR, or to more complicated structures, such as steroids, to explore cases where existing rules do not apply. Doing so might require the development of more sophisticated reordering techniques. Most importantly, the reordering concept utilizing laboratory data as a function of two or more controls can be applied to any molecular property. Analagous procedures apply as well to non-molecular based multi-compound material properties.

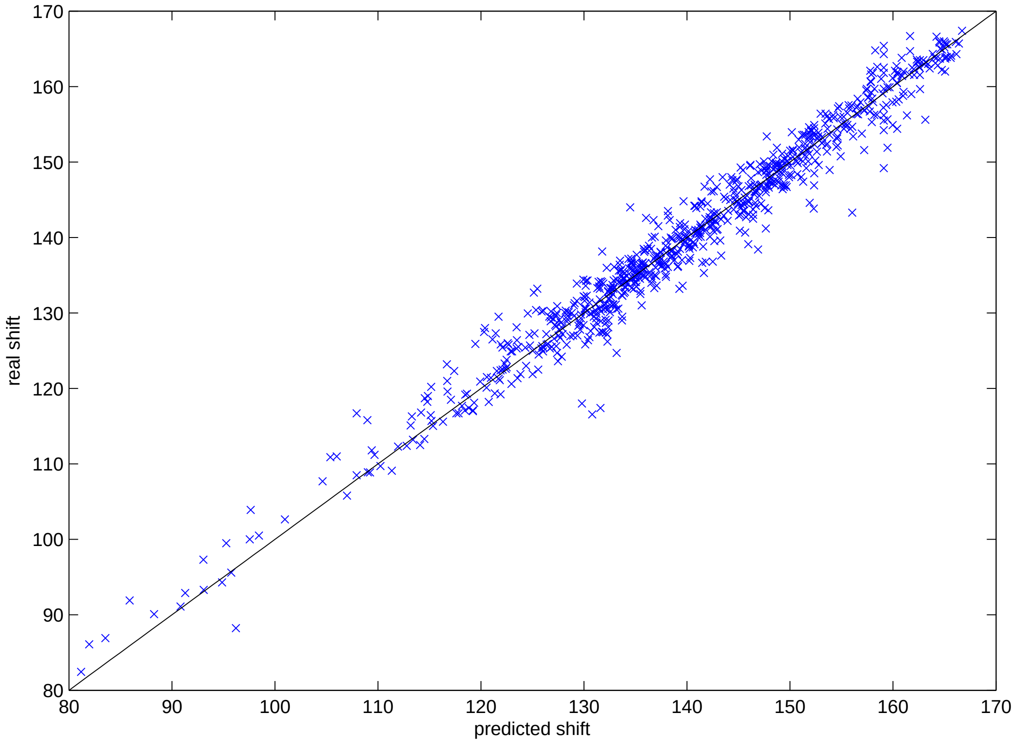

Figure 5: Truth plot for ^{13}C para-substituted benzene with $n_1 = 40$, $n_2 = 40$. A total of 1200 compounds were available with NMR shift data. Of those, 400 were used for training to deduce the proper ordering on the two sites. The remaining 800 were predicted by interpolating the landscape generated from the 400 compounds in the training set. The truth plot shows the excellent predictive capability of the reordering procedure.

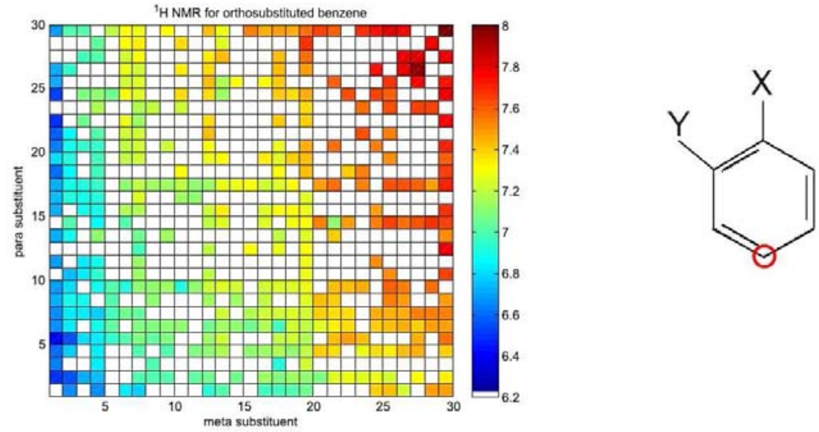

Figure 6: Reordered ^{1}H NMR data for orthosubstituted benzene, taken for a hydrogen on the circled carbon atom. The X position is "para" to the circled carbon, while the Y position is "meta" to the circled carbon.

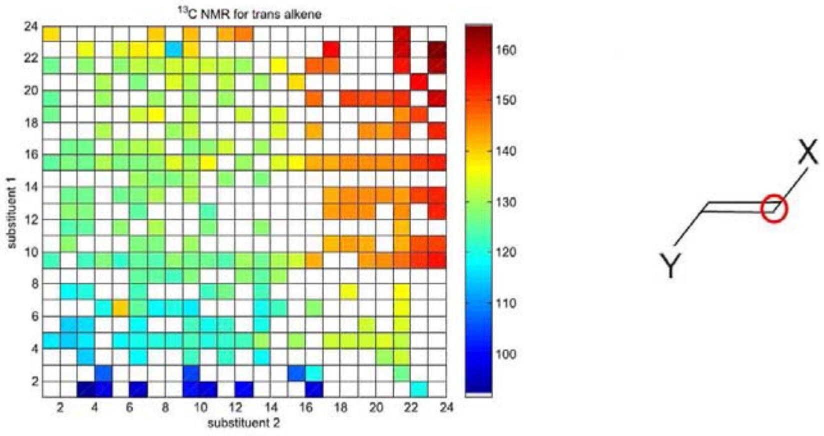

Figure 7: Reordered ^{13}C NMR data for trans-ethylene taken on the circled carbon atom. The X position is substituent 1, and the Y position substituent 2.

References

[1] Wu, R., Pechen, A., Rabitz, H., Hsieh, M., and Tsou, B. ”Control landscapes for observable preparation with open quantum systems”, *J. Math. Phys.* **49**, 022108 (2008).

[2] Boyd, Stephen and Vandenberghe, Lieven. *Convex Optimization*, Cambridge University Press, 2004.

[3] Schwabl, F. *Quantum Mechanics*, Third Edition, Springer-Verlag, New York, 2002.

[4] S. McAllister, et al, ”Descriptor-free molecular discovery in large libraries by adaptive substituent reordering”, *Bioorganic and Medicinal Chemistry Letters*, pp.5967-5970, 2008.

[5] Shenvi, N., Geremia, J. M., and Rabitz, H. ”Substituent Ordering and Interpolation in Molecular Library Optimization.” *J. Phys. Chem. A* **107**, 2066 (2003)