THE JOURNAL OF PHYSICAL CHEMISTRY A LETTERS

2006, *110*, 6391−6394
Published on Web 05/02/2006

General Method for the Dimension Reduction of Adaptive Control Experiments

Matthew A. Montgomery,[†] Robert R. Meglen,*[,‡] and Niels H. Damrauer*[,†]

Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309-0215, and Latent Structures LLC, 500 Manhattan Drive, Boulder, Colorado 80303

Received: February 23, 2006; In Final Form: April 14, 2006

Adaptive femtosecond control experiments are expanding the possibilities for using laser pulses as photophysical and photochemical reagents. However, because of the large number of variables necessary to perform these experiments (usually 100−200), it has proven difficult to elucidate the underlying control mechanisms from the optimized pulse shapes. If adaptive control is to become a widespread tool for examining chemical dynamics, methods must be developed that reveal latent control mechanisms. This manuscript presents a generally applicable method for dimension reduction of adaptive control experiments based on partial least squares regression analysis (PLS) of the normalized covariance matrix of the total data set. When applied to experimental results obtained in our laboratory, it shows that only seven fundamental dimensions from an original 208-dimension search space are needed to account for ∼90% of the variance in the observed fitness of 11 700 laser-pulse shapes explored during the optimization experiment. Furthermore, the seven dimensions have a remarkable regularity in their functional form. It is anticipated that this work will facilitate theoretical treatments directly linking the optimal fields to control mechanisms, allow quantitative comparisons of independent control results, and suggest new experimental methods for rapid adaptive searches.

The adaptive femtosecond control methodology first proposed by Rabitz[1] offers a powerful tool for chemical research by providing a means of manipulating complex molecular systems with light without specific prior knowledge of the Hamiltonian.[2] By coupling a broad-band laser-pulse shaper with iterative learning algorithms, researchers are able to search massive experimentally derived parameter spaces to discover laser-pulse shapes capable of inducing specific photochemical or photophysical outcomes. This has led to the discovery of control fields that would have been difficult or impossible to otherwise predict, including complex pulse shapes that achieve bond-selective chemistry,[3−8] that move energy in biological systems,[9,10] that enhance high-harmonic generation in gases,[11,12] and that selectively excite different chromophores in solution.[13−15] However, understanding how to infer control mechanisms from the adaptively discovered fields remains an outstanding question that must be addressed if the technique is to become a general tool for investigating chemical dynamics. As a step toward this goal, we present a generally applicable method for the dimension reduction of the variable space within these types of control problems. We stress that the methodology does not presuppose any specifics of the underlying control mechanism and that the experimental results are analyzed without arbitrarily biasing the search space. As a demonstration of this technique, we show that we can extract seven fundamental dimensions from an original 208 variable search space that account for ∼90% of the variance in the fitness of 11 700 pulse shapes explored in our laboratory. Herein, fitness refers to the quality of any laser pulse explored by the adaptive algorithm with respect to the user-defined optimization goal. Only three dimensions are needed to account for 82% of the variance. Furthermore, the seven dimensions have remarkable regularity in their functional form, which should facilitate future theoretical treatments linking the optimal fields to control mechanisms.

In many adaptive control experiments to date, astronomically large search spaces are made possible by manipulating the frequency-dependent phase of broad-band laser pulses using many (on the order of 100) independent phase parameters. In some reports, the adaptively discovered fields reveal regular structure such as evenly spaced pulse-like features.[5,9] In these cases, mechanistic insight has been gained using parametrization techniques that recreate salient pulse characteristics or pump−probe experiments that test the adaptively discovered features. However, there is no guarantee that complex control fields will have readily interpretable structure. Often, the massive parameter space that facilitates control in the first place also serves to obscure mechanistic insight because the heuristic optimization algorithms do not distinguish between pulse features necessary for control and those that contribute negligibly to fitness.[16,17] In these cases, there heretofore exists no rational basis to parametrize or reduce the search space (such as searching in orders of a Taylor expansion of the chirp) from the discovered solution. In fact, we would argue that an experimenter is apt to further obfuscate the latent control mechanism by implementing an arbitrary parametrization or search space reduction even if control can be achieved.

Nonetheless, researchers must be able to understand pulse shapes in the context of an intellectually tractable number of variables if control mechanisms are to be investigated. In

* Corresponding authors. E-mail: N.H.D., niels.damrauer@colorado.edu; R.R.M., bmeglen@latentstructures.com.
[†] University of Colorado at Boulder.
[‡] Latent Structures LLC.

10.1021/jp061160l CCC: $33.50

essence, adaptively shaped laser pulses must be distilled to their essential structure. Much recent research has been directed toward dimension reduction of adaptive control experiments. Bartels et al. have analyzed pulse statistics to corroborate predicted control mechanisms.[12] Weinacht's group has shown that the dimensionality of the search space can be significantly reduced through the appropriate selection of a basis set and that, in principle, linear transformations of the search space can be used for this purpose.[17−19] Bucksbaum's group has applied multivariate statistical methods to their adaptive results to make such transformations and conclude that diagonalization of the covariance matrix is an appropriate means of extracting the principal control directions of the system.[20]

Our analysis suggests that statistical methods can pay too much attention to the mechanism of the search algorithm rather than mechanisms of control. It is crucial to quantify the correlation of phase-parameter variance with experimentally measured pulse fitness. Furthermore, in creating new basis functions, it is essential that discontinuities in phase functions are properly accounted for prior to the statistical treatments.

The experimental setup we have used (see Supporting Information) is similar to one previously described by Gerber's group.[14] Briefly, broadband laser pulses ($\sim$800 $\pm$ 17 nm) are shaped using "phase functions" composed of 208 independent variables across the laser-pulse spectrum. The number of pixels is chosen by inspection to include greater than 95% of spectral intensity as detected on an Ocean Optics SD2000 spectrometer. An immense number of light fields is possible ($>10^{400}$), each with a different time ordering of its frequency components. The output of the pulse shaping device is split into two pulse trains. One impinges on a 298 K sample of $[Ru(dpb)_3](PF_6)_2$ in acetonitrile (where dpb = 4,4′-diphenyl-2,2′-bipyridine). The linear absorption spectrum of this system has negligible absorbance at the wavelengths contained in the laser pulse, but electronic excitation occurs if the molecule absorbs two or more photons from the shaped field. The relative multiphoton excitation efficiency is monitored with a 640 $\pm$ 5 nm spontaneous emission signal from the thermalized triplet metal-to-ligand charge transfer state. The second pulse train is passed through a thin BBO crystal to generate second harmonic (SHG) of the fundamental. This signal reports the relative intensity of each laser pulse tested by the adaptive algorithm. Gerber's group has shown that maximizing the *ratio* of these two signals (emission/SHG) removes the two-photon intensity dependence of either separate physical process. This allows the adaptive algorithm to take advantage of molecule-specific information to discover optimal pulse shapes.[14] For the data analyzed herein, the algorithm was run for 195 generations with 60 laser pulses per generation (11700 total pulses). The optimal pulses in this experiment approximately doubled the emission/SHG ratio with respect to randomly encoded pulse shapes. We have chosen this experiment for the development and application of our statistical methods because the optimal pulse shapes show complex time and frequency-dependent structure as reported elsewhere.[14]

To begin deconstructing patterns that emerge in optimal pulse shapes, we have developed a general analysis tool that is based on partial least squares regression (PLS) of the normalized covariance matrix.[21,22] This is a soft-modeling technique commonly used in chemometrics to understand multiconstituent spectra. Conceptually, the regression algorithm accomplishes three tasks. First, the phase variables are fit to a minimum dimensional hyperplane in the search space. The experimental data points are then projected onto this plane to determine the variation of fitness relative to position on the plane. Finally,

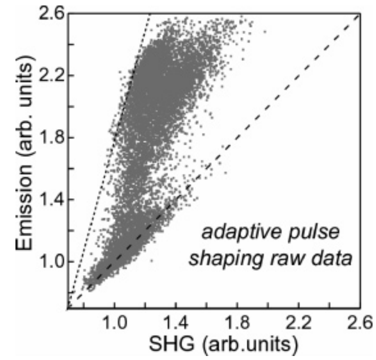

Figure 1. Emission versus SHG for all 11 700 laser pulses explored during the optimization experiment.

the plane is expressed with an orthogonal basis set such that the first basis vector points along the direction of greatest fitness variance on the hyperplane. The second basis vector points along the direction of next greatest variance and so forth. These orthogonal vectors are linear combinations of the original 208 phase variables. For further description and discussion, see the Supporting Information.

We have applied our method to a data set consisting of 11 700 spectral phase functions explored in a single 195-generation optimization experiment. It is possible to represent the full data set as a plot of the measured molecular emission versus second harmonic for each pulse (Figure 1). Because the algorithm is trying to maximize the ratio of these two signals, the best pulses are those that lie in the upper left of the graph, on or above the dotted line. Randomly encoded phase functions that are used to start the algorithm, as well as near bandwidth limited pulses not shown here,[14] lie on or near the dashed line with a ratio of 1 (arbitrary units).

The adaptive algorithm manipulates the phase modulo 2π and the resultant phase functions for all but the simplest laser fields can exhibit discontinuities between adjacent pixels. For modeling purposes this is problematic because the relative phase, rather than the absolute phase, determines pulse fitness. Thus, it is essential that the functions be made continuous prior to PLS analysis. We implement an algorithm that "unwraps" each function such that there are no phase discontinuities between adjacent pixels greater than π. We note in agreement with one of our reviewers that unwrapping is only necessary if the phase has been searched modulo 2π. Some pulse shaping techniques, such as those that use acousto-optic modulators or deformable mirrors, do not produce discontinuous phase functions. However, in either case the PLS analysis can implemented. The normalized covariance matrix can then be calculated and diagonalized to reveal the fundamental directions along which phase functions vary in the total data set. These directions are known as principal components (PCs) and they are ranked by the percentage of the total phase variance that they embody. White et al. have argued that the few PCs of statistical importance (highest percentage of the variance) reveal the fundamental degrees of freedom in the control Hamiltonian.[20] Although variance does contain information about the fitness, their relationship is necessarily obscured by the heuristic nature of the evolutionary algorithm's search mechanics.

Here we report that it is imperative to directly correlate phase data with laser-pulse fitness. The PLS analysis regresses the normalized covariance matrix of the 11 700 phase functions against the variance of the fitness for the pulses. This reveals the fundamental directions wherein changes to the phase functions correlate most strongly to fitness. The extracted vectors

TABLE 1: Phase and Fitness Variance for PLS Dimensions

PLS dimensions	% variance in phase	% variance in fitness
PC 1	62	47
PC 2	31	24
PC 3	1	11
PC 4	2	2
PC 5	1	2
PC 6	< 1	2
PC 7	< 1	1

are quantified in terms of both the percentage variance in phase and fitness. As shown in Table 1, seven orthogonal interpixel phase relationships account for 89% of the observed fitness variance. It is important to note that the third PC accounts for 11% of the fitness variance despite exhibiting negligible phase variance. This would have been overlooked had we only considered the diagonalization of the covariance matrix.

Each of these basis vectors can be plotted in terms of relative phase versus pixel number. If the phase unwrapping procedure is neglected prior to the PLS analysis, highly discontinuous functions are discovered. For purposes of comparison, the incorrect PC1 is shown here in Figure 2. When phase unwrapping is employed prior to PLS analysis, the extracted basis functions bear remarkable simplicity and qualitative similarity to certain orthogonal functions of mathematical physics such as particle-in-a-box wave functions. The first seven basis functions are shown in Figure 3.

The small number of simple functions needed to account for the majority of the fitness reveals that highly structured pulse features are not necessary for increasing emission/SHG in this experimental system. This is consistent with a control mechanism that exploits broad features in the two-photon absorption spectrum of the molecule, as suggested by the Gerber group[14] and confirmed by the Joffre group for a related experiment using the organic chromophore Coumarin 460.[15] Most importantly, the analysis reveals in a completely general and unbiased fashion which and how many independent experimental "knobs" are needed to manipulate control mechanisms unveiled by the larger adaptive search. We are now conducting optimizations with far fewer control variables consisting of the basis vectors revealed by PLS analysis as well as actual particle-in-a-box wave functions. Such adaptive searches are accomplished in a fraction of the time necessary for a 208-parameter optimization. We are also exploring how different linear combinations of these eight basis functions manifest themselves in the optimal control fields. Finally, we are applying these general analysis methods to new

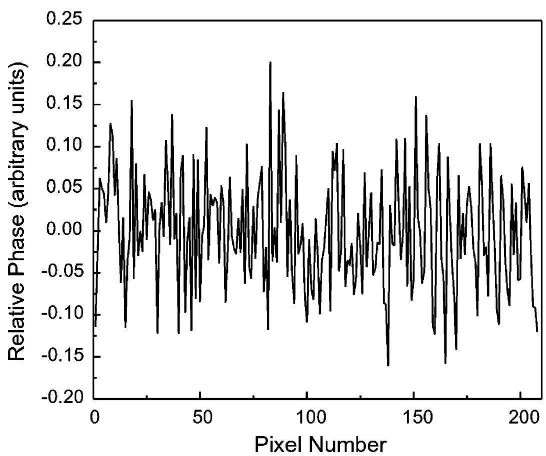

Figure 2. First PC if phase unwrapping is *neglected* prior to the PLS analysis. This highly complex function should be compared with PC 1 from Figure 3.

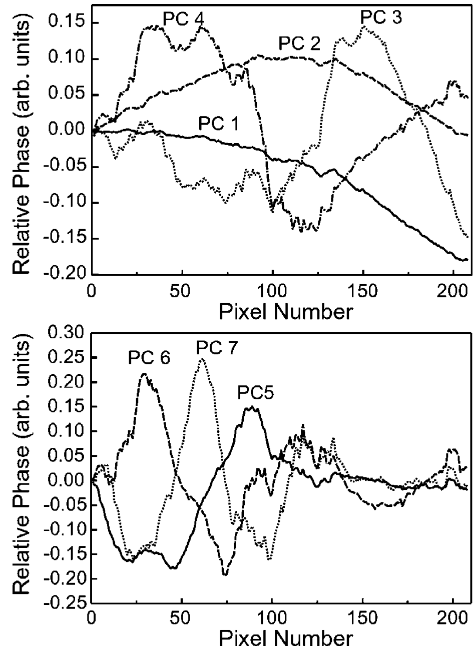

Figure 3. Plots showing PC's 1−4 (top) and 5−7 (bottom). These reveal the interpixel phase relationships that account for 89% of the fitness variance in the 11 700-pulse data set.

control results discovered in our laboratory. This is bringing us much closer to being able to read such distributions as a record in time and frequency of control mechanisms uncovered by adaptive experiments.

Acknowledgment. This work was supported by the University of Colorado and the ACS Petroleum Research Fund.

Supporting Information Available: Laboratory and analysis methods. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Judson, R. S.; Rabitz, H. *Phys. Rev. Lett.* **1992**, *68* (10), 1500−1503.

(2) Rabitz, H. A.; Hsieh, M. M.; Rosenthal, C. M. *Science* **2004**, *303* (5666), 1998−2001.

(3) Assion, A.; Baumert, T.; Bergt, M.; Brixner, T.; Kiefer, B.; Seyfried, V.; Strehle, M.; Gerber, G. *Science* **1998**, *282* (5390), 919−922.

(4) Levis, R. J.; Menkir, G. M.; Rabitz, H. *Science* **2001**, *292*, 709−713.

(5) Daniel, C.; Full, J.; Gonzalez, L.; Lupulescu, C.; Manz, J.; Merli, A.; Vajda, S.; Woste, L. *Science* **2003**, *299* (5606), 536−539.

(6) Cardoza, D.; Baertschy, M.; Weinacht, T. *Chem. Phys. Lett.* **2005**, *411* (4−6), 311−315.

(7) Cardoza, D.; Baertschy, M.; Weinacht, T. *J. Chem. Phys.* **2005**, *123* (7), 074315.

(8) Damrauer, N. D.; Dietle, C.; Krampert, G.; Lee, S. H.; Jung, K. H.; Gerber, G. *Eur. Phys. J. D* **2002**, *20*, 71−76.

(9) Herek, J. L.; Wohlleben, W.; Cogdell, R. J.; Zeidler, D.; Motzkus, M. *Nature* **2002**, *417* (6888), 533−535.

(10) Wohlleben, W.; Buckup, T.; Herek, J. L.; Motzkus, M. *Chem. Phys. Chem.* **2005**, *6* (5), 850−857.

(11) Bartels, R.; Backus, S.; Zeek, E.; Misoguti, L.; Vdovin, G.; Christov, I. P.; Murnane, M. M.; Kapteyn, H. C. *Nature* **2000**, *406* (6792), 164−166.

(12) Bartels, R. A.; Murnane, M. M.; Kapteyn, H. C.; Christov, I.; Rabitz, H. *Phys. Rev. A* **2004**, *70* (4), 1−5.

(13) Brixner, T.; Damrauer, N. H.; Niklaus, P.; Gerber, G. *Nature* **2001**, *414* (6859), 57−60.

(14) Brixner, T.; Damrauer, N. H.; Kiefer, B.; Gerber, G. *J. Chem. Phys.* **2003**, *118*, 3692−3701.

(15) Ogilvie, J. P.; Kubarych, K. J.; Alexandrou, A.; Joffre, M. *Opt. Lett.* **2005**, *30* (8), 911−913.

(16) Geremia, J. M.; Zhu, W. S.; Rabitz, H. *J. Chem. Phys.* **2000**, *113* (24), 10841−10848.

(17) Cardoza, D.; Trallero-Herrero, C.; Langhojer, F.; Rabitz, H.; Weinacht, T. *J. Chem. Phys.* **2005**, *122* (12), 124306.
(18) Langhojer, F.; Cardoza, D.; Baertschy, M.; Weinacht, T. *J. Chem. Phys.* **2005**, *122* (1), 014102.
(19) Cardoza, D.; Langhojer, F.; Trallero-Herrero, C.; Monti, O. L. A.; Weinacht, T. *Phys. Rev. A* **2004**, *70* (5), 053406.
(20) White, J. L.; Pearson, B. J.; Bucksbaum, P. H. *J. Phys. B: At., Mol. Opt. Phys.* **2004**, *37* (24), L399−L405.
(21) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J. *Siam J. Sci. Stat. Comput.* **1984**, *5* (3), 735−743.
(22) Wold, S.; Sjostrom, M.; Eriksson, L. *Chemometrics Intelligent Lab. Syst.* **2001**, *58* (2), 109−130.