Gradient algorithm applied to laboratory quantum control

Jonathan Roslund* and Herschel Rabitz
Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA
(Received 19 February 2009; published 18 May 2009)

The exploration of a quantum control landscape, which is the physical observable as a function of the control variables, is fundamental for understanding the ability to perform observable optimization in the laboratory. For high control variable dimensions, trajectory-based methods provide a means for performing such systematic explorations by exploiting the measured gradient of the observable with respect to the control variables. This paper presents a practical, robust, easily implemented statistical method for obtaining the gradient on a general quantum control landscape in the presence of noise. In order to demonstrate the method's utility, the experimentally measured gradient is utilized as input in steepest-ascent trajectories on the landscapes of three model quantum control problems: spectrally filtered and integrated second harmonic generation as well as excitation of atomic rubidium. The gradient algorithm achieves efficiency gains of up to approximately three times that of the standard genetic algorithm and, as such, is a promising tool for meeting quantum control optimization goals as well as landscape analyses. The landscape trajectories directed by the gradient should aid in the continued investigation and understanding of controlled quantum phenomena.

DOI: 10.1103/PhysRevA.79.053417 PACS number(s): 32.80.Qk

I. INTRODUCTION

The implementation of quantum control in the laboratory has been facilitated through the confluence of several technologies, including high repetition rate ultrafast laser sources [1], femtosecond pulse shaping technology [2], and practical high-dimensional learning algorithms [3]. In particular, the field has effectively employed learning algorithms in a closed-loop fashion to efficiently search the typically high-dimensional control spaces intrinsic to these problems [4].

Evolutionary algorithms were originally believed to be well suited for the optimization of quantum control problems because their global search capability was thought to be necessary for locating a lone solution in an exponentially large search space. In addition, the algorithm's stochastic nature was also believed to be important for escaping from any local suboptimal maxima on the underlying search landscape and handling noise that is inevitably present in a laboratory setting. Indeed, the mounting successes of evolutionary algorithm-guided optimizations across a broad range of problems [5–7] have led to their adoption as a standard tool for experimental quantum control.

This general success, as well as the relative ease of quickly locating a family of high-quality robust solutions, prompted analyses of the underlying quantum control landscape, which is defined as the physical observable as a function of the control field. Intuition had originally surmised that the highly nonlinear mapping between control variables and the measured observable would result in an extremely complex, structurally rich landscape upon which exploration would best be handled with the machinery of stochastic algorithms. However, recent theoretical analyses have begun to give insight into the true nature and topology of these search landscapes [8,9]. In particular, it has been shown that the landscapes of general quantum observables are not infused with suboptimal local maxima and minima as originally believed but rather only have global maxima (minima) at the highest (lowest) attainable value of the observable [10,11]; in certain cases, nontrapping saddle structures may also be present. Additionally, quantum control problems exhibit a large solution multiplicity, and repeated optimizations will often converge to several equally successful unique control solutions [12,13].

With the current knowledge of the underlying search landscape topology, a natural question is whether an optimal algorithm exists that is best suited for maneuvering quantum control landscapes. The stochastic nature of the genetic algorithm (GA) or other evolutionary algorithms is still attractive for locating high-quality solutions in the presence of laboratory noise. However, given the existence of a large solution multiplicity and the monotonically increasing nature of the control landscapes, it is possible that a local search algorithm is better suited in terms of both efficiency and uncovered landscape information rather than performing a global search driven by a stochastic-based algorithm. Although the generic *topology* of unconstrained quantum control landscapes is surprisingly simple, little is known about the landscape local *structure*, which is likely system dependent. Naturally, knowledge of a problem's local structure would be advantageous for uncovering a global nonlinear transformation to reveal the "proper" curvilinear control variables and possibly provide mechanistic insights.

One such local algorithm that is particularly amenable to the topology of unconstrained quantum control landscapes is gradient ascent. Regardless of the landscape local structural complexity, the monotonicity of quantum control landscapes guarantees a successful optimization when operating with low noise and adequately unconstrained controls. In addition to providing the direction of steepest ascent, the gradient provides myopic information about control landscape features. Excursions over arbitrarily high-dimensional landscapes are only really practical with trajectory-based methods, and knowledge of the gradient provides an entree into trajectory-based exploratory techniques, such as observable

*jroslund@princeton.edu

tracking [12,14] and Pareto front exploration [15,16], to not only climb but also traverse the landscape at will in any direction and report on identified local features.

It should be noted that trap free landscapes can be envisioned in which a global search algorithm would still be better suited in terms of efficiency than local algorithms, including gradient ascent. One such example is a model landscape described as a high pitch helical structure. Although gradient ascent would likely be less efficient than global optimization in this case, the curvilinear control field coordinate transformations intrinsic to the structure of the landscape would likely only be discovered with a local algorithm. Thus, even for problems in which the gradient may not be efficient, it is capable of providing detailed knowledge about a given problem's underlying local landscape structure.

An outstanding subject concerns the alteration of the generic quantum control landscape topology in the presence of constraints (e.g., fixed center wavelength, limited bandwidth and fluence, and finite pixel resolution), which are inevitably present in any laboratory situation [17]. Thus, invariably, experimental landscapes will not exhibit perfect monotonicity. Yet, implementation of the gradient algorithm is also of interest here since suboptimal convergence indicates a presence of false traps, and gradient information may reveal the source of the trap (e.g., improper control field basis, insufficient energy, etc.) and possible means for escape.

Here we develop a practical laboratory methodology for obtaining the gradient of an observable with respect to the control variables in the presence of noise. Rather than obtaining the gradient via finite differences, which is known to suffer inaccuracies due to the presence of noise, a statistical estimation of the gradient is obtained. A laboratory-based steepest-ascent algorithm based on this statistically determined gradient is then implemented for the optimization of several model quantum control problems, and its performance is evaluated in reference to that of a standard GA.

The paper is organized as follows. Section II outlines the methodology for statistically estimating the gradient in the laboratory. The nature of this gradient is then explored with a simple mathematical model and after briefly outlining the experimental details in Sec. III, its accuracy and convergence properties are examined for several quantum control systems in Sec. IV. A steepest-ascent algorithm based on this statistical gradient is then implemented for three prototypical quantum systems in Sec. V. Finally, a summary of the results and outlook toward potential applications is presented in Sec. VI.

II. DETERMINATION OF THE STATISTICAL GRADIENT

Obtaining a high-quality gradient in the presence of laboratory noise is a challenging task since the standard finite difference technique is especially susceptible to field fluctuations and data errors. Indeed, determination of a gradient based on finite differences of experimental data often leads to noise amplification thereby producing inadequate or spurious results. This problem may be somewhat diminished by first smoothing or filtering the noise, but the results can be quite varied. Instead, a statistical, robust method for derivative estimation is formulated here based on taking first moments of the data, which are easily measured in the laboratory.

Consider the first moment of a continuous, differentiable, multidimensional landscape $J(\vec{x})$ function of the control variables $\vec{x}=(x_1,x_2,\ldots,x_D)$ with $-\infty\leq x_i\leq\infty$, where the expectation value is performed over a multidimensional probability distribution $P(\vec{x})$ centered about the origin $\vec{x}=0$, i.e.,

$$\langle x_i J(\vec{x}_0+\vec{x})\rangle = \int_{-\infty}^{\infty} x_i J(\vec{x}_0+\vec{x})P(\vec{x})d\vec{x}, \tag{1}$$

where $d\vec{x}=dx_1dx_2\ldots dx_D$. Assuming that the function $J(\vec{x})$ may be Taylor-expanded around the point $\vec{x}_0$ gives

$$J(\vec{x}_0+\vec{x}) \simeq J(\vec{x}_0)+\sum_j^D \left.\frac{\partial J}{\partial x_j}\right|_{\vec{x}_0} x_j+\frac{1}{2}\sum_{j,k}^D \left.\frac{\partial^2 J}{\partial x_j\,\partial x_k}\right|_{\vec{x}_0} x_j x_k + \frac{1}{6}\sum_{j,k,l}^D \left.\frac{\partial^3 J}{\partial x_j\,\partial x_k\,\partial x_l}\right|_{\vec{x}_0} x_j x_k x_l + \cdots. \tag{2}$$

For small deviations $\vec{x}$ around the point $\vec{x}_0$ [i.e., considering that $P(\vec{x})$ is a sufficiently narrow distribution], this expansion may be substituted into Eq. (1) to obtain

$$\begin{aligned}\langle x_i J(\vec{x}_0+\vec{x})\rangle \simeq\ & J(\vec{x}_0)\int_{-\infty}^{\infty} x_i P(\vec{x})d\vec{x} \\ &+\sum_j^D \left.\frac{\partial J}{\partial x_j}\right|_{\vec{x}_0}\int_{-\infty}^{\infty} x_i x_j P(\vec{x})d\vec{x} \\ &+\frac{1}{2}\sum_{j,k}^D \left.\frac{\partial^2 J}{\partial x_j\,\partial x_k}\right|_{\vec{x}_0}\int_{-\infty}^{\infty} x_i x_j x_k P(\vec{x})d\vec{x} \\ &+\frac{1}{6}\sum_{j,k,l}^D \left.\frac{\partial^3 J}{\partial x_j\,\partial x_k\,\partial x_l}\right|_{\vec{x}_0}\int_{-\infty}^{\infty} x_i x_j x_k x_l P(\vec{x})d\vec{x} \\ &+\cdots.\end{aligned} \tag{3}$$

It is now assumed that the multidimensional probability distribution function $P(\vec{x})$ is separable and may be written as the product of D one-dimensional uncorrelated, symmetric probability distributions $p(x_i)$, i.e., $P(\vec{x})=p(x_1)p(x_2)\ldots p(x_D)$, where $p(x_i)=p(-x_i)$. While the D probability distributions need not be identical, it is assumed here for simplicity. With these assumptions, Eq. (1) becomes

$$\langle x_i J(\vec{x}_0+\vec{x})\rangle \simeq \mu_2\frac{\partial J(\vec{x}_0)}{\partial x_i}+\frac{1}{6}\left(\mu_4\frac{\partial^3 J(\vec{x}_0)}{\partial x_i^3}+3\mu_2^2\sum_{j\neq i}^D \frac{\partial^3 J(\vec{x}_0)}{\partial x_i\,\partial x_j^2}\right) + \cdots, \tag{4}$$

where μ_2 and μ_4 are the second and fourth moments of the probability distribution function $p(x)$, respectively. This equation may be rearranged to yield

$$\frac{\partial J(\vec{x}_0)}{\partial x_i} \simeq \frac{1}{\mu_2}\int_{-\infty}^{\infty} x_i J(\vec{x}_0+\vec{x})P(\vec{x})d\vec{x} - \frac{1}{6}\left(\frac{\mu_4}{\mu_2}\frac{\partial^3 J(\vec{x}_0)}{\partial x_i^3} + 3\mu_2\sum_{j\neq i}^{D}\frac{\partial^3 J(\vec{x}_0)}{\partial x_i\,\partial x_j^2}\right) + \cdots . \quad (5)$$

Note that no assumption has been made as to the nature of the function $J(\vec{x})$ or form of the Hessian, i.e., $\frac{\partial^2 J}{\partial x_i \partial x_j}$, at the point $\vec{x}_0$. By choosing a symmetric probability distribution function $p(x_i)=p(-x_i)$, there are no contributions in Eq. (5) from the first- and third-moment terms in the expansion of Eq. (3).

The analysis above shows that the gradient $\vec{\nabla}J(\vec{x})|_{\vec{x}_0}$ of the landscape function $J(\vec{x})$ at a point $\vec{x}_0$ can be estimated from the first moment $\langle x_i J(\vec{x}_0+\vec{x})\rangle$ through locally sampling the neighborhood of the point $\vec{x}_0$. The only sampling requirements are that the variables x_i be drawn from statistically independent and symmetric distributions. The error implicit to this statistically estimated gradient may then be written as

$$\frac{\langle x_i J(\vec{x}_0+\vec{x})\rangle}{\mu_2} - \frac{\partial J(\vec{x}_0)}{\partial x_i} \simeq \frac{\mu_4}{\mu_2}\epsilon_{i,1} + \mu_2\epsilon_{i,2} + \cdots , \quad (6)$$

where $\epsilon_{i,1}$ and $\epsilon_{i,2}$ are comprised of the third derivative terms on the right-hand side of Eq. (5). Thus, for a narrow sampling distribution ($\mu_2 \to 0$ and $\mu_4/\mu_2 \to 0$), the statistical gradient approaches the true gradient. Consequently, for a modest choice of the sampling domain on a slowly varying landscape, the gradient may be well approximated by the weighted first moment.

In practice, the present work will estimate the gradient from the sampling of N laboratory data points $\vec{x}_n$ according to

$$\frac{\partial J(x_0)}{\partial x_i} \simeq \frac{1}{\mu_2 N}\sum_{n}^{N} x_{i,n} J(\vec{x}_0+\vec{x}_n), \quad (7)$$

where the variables x_n are drawn from a Gaussian probability density with a standard deviation of σ, i.e., $p(x_n) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp[-\frac{x_n^2}{2\sigma^2}]$ and the pertinent central moments are $\mu_2 = \sigma^2$ and $\mu_4 = 3\sigma^4$. Other suitable symmetric distributions could just as well be employed to estimate the gradient.

In order to illustrate the underlying principles of this method before considering its performance in the laboratory, the statistical gradient is evaluated with an analytical, two-dimensional, nonseparable model landscape,

$$J(x,y) = \exp[-(x^2+y^2+xy)]. \quad (8)$$

The true partial derivative $\frac{\partial J(x,y)}{\partial x}$ evaluated at the reference point (x_0,y_0) is

$$\left.\frac{\partial J(x,y)}{\partial x}\right|_{(x_0,y_0)} = -[y_0+2x_0]\exp[-(x_0^2+y_0^2+x_0y_0)], \quad (9)$$

and the weighted first moment at the point (x_0,y_0) may be exactly calculated,

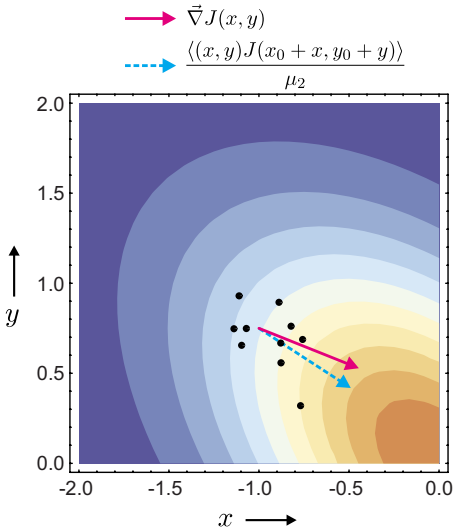

FIG. 1. (Color online) Evaluation of the statistical gradient estimator on a model two-dimensional landscape. The gradient is estimated (blue dashed vector) at the point $(x_0,y_0)=(-1.0,0.75)$. The local sampling is drawn from a Gaussian distribution with a width of $\sigma=0.15$ and a modest $N=10$ points. For comparison, the analytical gradient is shown (red solid vector).

$$\frac{\langle xJ(x_0+x,y_0+y)\rangle}{\mu_2} = -\frac{[y_0+(2+3\sigma^2)x_0]}{f(\sigma)} \times \exp\left[-\frac{(x_0^2+y_0^2)(1+1.5\sigma^2)+x_0y_0}{1+4\sigma^2+3\sigma^4}\right], \quad (10)$$

where $f(\sigma)$ is a function of σ with $\lim_{\sigma\to 0} f(\sigma)=1$ [18]. This weighted first moment reduces to the true gradient when $\sigma \to 0$. Importantly, the estimated derivative does not effectively alter the intrinsic topology of the landscape by introducing or removing couplings between the canonical variables as may readily be seen by examining the locally averaged landscape $\tilde{J}(x_0,y_0)=\langle J(x_0+x,y_0+y)\rangle$. However, the local structure is affected since variable couplings are diminished with increasing σ, which effectively transforms the model ellipsoid into a more circular form.

As an example of statistically determining the gradient, the estimator is evaluated according to Eq. (7) for $N=10$ points drawn from a Gaussian distribution ($\sigma=0.15$) about the point $(x_0,y_0)=(-1.0,0.75)$ for the landscape described by Eq. (8). The strict variate sampling symmetry required for analytically zero odd-moment terms is abandoned in favor of increased landscape sampling for a fixed sample size N. Nonetheless, even with this very modest sampling, Fig. 1 shows that the recovered gradient faithfully approximates the exact analytical derivative. A step in the direction of this statistically estimated gradient results in a clear yield increase.

III. EXPERIMENTAL DETAILS

A commercial Ti:sapphire femtosecond laser system produces amplified pulses centered at 797 nm with a band-

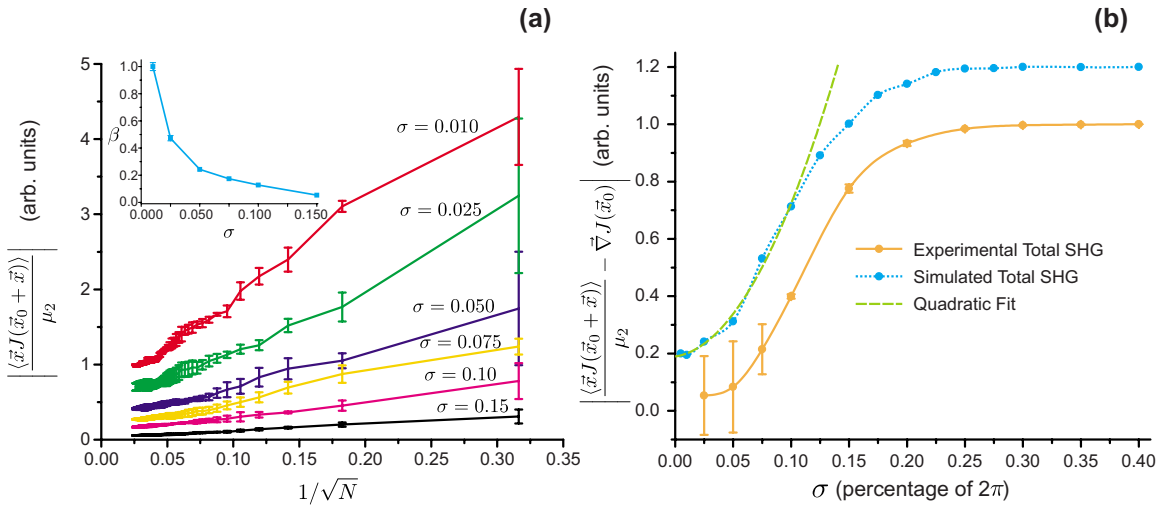

FIG. 2. (Color online) Experimentally measured (a) statistical gradient convergence rates and (b) accuracy for $D=128$ filtered and total SHG, respectively. Panel (a) displays the $1/\sqrt{N}$ convergence rate for various values of the sampling distribution width σ for filtered SHG. This rate is readily fit to the form $\alpha+\beta/\sqrt{N}$, and the retrieved β values, which describe the sensitivity of the estimated gradient to sample size N, are shown in the inset. Panel (b) displays the gradient norm deviation for total SHG computed with $N=1500$ samples for various values of the sampling distribution width σ. The computed norm and quadratic fit about $\sigma=0$ for the gradient on an equivalent simulated noise-free landscape (vertically offset for visualization) is shown for reference. Each curve is normalized to the computed norm deviation at large σ. The shapes of both the experimental and simulated curves agree with the form predicted by Eq. (11) as $\mu_2 \sim \sigma^2$.

width of $\Delta\lambda \sim 10$ nm, which gives pulses of $\Delta\tau \sim 100$ fs duration [full width at half maximum (FWHM)]. These pulses are delivered to a $4-f$ configuration pulse shaper with a programmable 128 pixel liquid-crystal spatial light modulator (SLM-256, CRI) for phase-only modulation. Three control systems are studied with the gradient algorithm.

First, the total second harmonic generation (SHG) signal is observed by focusing amplified pulses onto a 100 μm type-I β-BaB_2O_4 (BBO) crystal, and the spectrally integrated SHG signal is recorded with a photodiode and boxcar integrator [13]. For the second case of filtered SHG, unamplified seed pulses are focused onto a 100 μm type-I BBO crystal, and the resultant upconverted light is analyzed with a spectrometer [19]. The filtered SHG signal is recorded at 398.35 nm (i.e., $\lambda_0/2$) with a spectral window of $\Delta\lambda \sim 0.06$ nm. The 80 MHz repetition rate of the oscillator guarantees a high signal-to-noise ratio. For the third system of atomic rubidium, amplified pulses with an energy of 135 μJ are delivered to a 75 mm vapor cell containing atomic rubidium maintained at 100 °C. The shaped pulses induce $5S \rightarrow 5P \rightarrow 5D$ transitions [20]. The atomic $5D$ state decays radiatively to the $6P$ state, which then undergoes further radiative decay to the ground state by emitting an observed 420 nm photon. No selection is made based on spin-orbit coupling, and the visible $6P \rightarrow 5S$ fluorescence is collected in a direction orthogonal to the incident laser beam.

IV. STATISTICAL GRADIENT ACCURACY AND CONVERGENCE

Prior to implementing a gradient-based laboratory algorithm to climb the landscapes, a thorough understanding is necessary with regard to the accuracy and convergence properties of evaluating the statistical gradient. Since the statistical gradient is based on a weighted first moment, its rate of convergence is expected to scale as $1/\sqrt{N}$, where N is the number of locally sampled points on the control landscape [21]. Additionally, the accuracy of the technique is assessed by monitoring the gradient norm as the standard deviation σ of the landscape sampling distribution is increased. The expected accuracy is understood by computing a norm from Eq. (6) (where it is assumed here for simplicity that $\mu_4 = 3\mu_2^2$ in keeping with the use of Gaussian sampling distributions),

$$\left| \frac{\langle \vec{x} J(\vec{x}_0 + \vec{x}) \rangle}{\mu_2} - \vec{\nabla} J(\vec{x}_0) \right| \simeq \mu_2 |\vec{\epsilon}|, \tag{11}$$

where nonlinear terms in μ_2 are discarded and $|\vec{\epsilon}|$ includes third derivative terms. Thus, the norm of the statistical gradient deviation from the true gradient is expected to vary quadratically with the local sampling domain σ (i.e., $\mu_2 \sim \sigma^2$).

The convergence is carefully assessed by experimentally measuring the gradient at a fixed location on the quantum control landscape for filtered SHG [22]. A negative quadratic phase is applied to the pulse shaper in order to obtain a yield of $S_f \sim 0.5$ [$\phi(\Omega) \sim -4.5 \cdot 10^3$ fs^2 Ω^2]. The underlying quantum control landscape is then locally sampled about $\phi(\Omega)$ in an unrestricted fashion with points drawn from a Gaussian distribution.

The measured convergence properties for the statistical gradient are shown in Fig. 2(a). The $D=128$ dimensional gradient is estimated by Eq. (7), where the sampling width σ of a phase variable x_i is a fraction of the total spectral phase domain $\phi(\omega_i) \in [0, 2\pi]$ (e.g., $\sigma=0.1$ corresponds to a spectral phase space sampling width of 0.63). There is no loss of generality in defining the spectral phase domain to be $[0, 2\pi]$

as integer multiples of 2π may always be added to the control variable without altering the physical control field; therefore, constraint of the shaper pixels to lie in this domain does not affect convergence of the integrals defined over $-\infty \leq x_i \leq \infty$. The experimentally computed gradient norms exhibit the expected $1/\sqrt{N}$ convergence for all tested values of σ. These norms also agree with the trends expected from Eq. (11) in that the norm varies nonlinearly with σ. Each of these convergence curves may be fit to the form $\left|\frac{\langle \vec{x}J(\vec{x}_0+\vec{x})\rangle}{\mu_2}\right| = \alpha + \frac{\beta}{\sqrt{N}}$. The fit parameter β describes the sensitivity of the computed norm to the number of sampled data points N, and the retrieved values are shown in the inset of Fig. 2(a). For the purpose of optimization, a high-quality gradient is desired while also operating with minimal sampling. A good compromise between these two conflicting objectives is obtained in this case with a sampling width of $\sigma=0.05$.

In order to establish the accuracy of the method, the recovered statistical norms calculated with a fixed sample size of $N=1500$ for total SHG ($D=128$) are shown in Fig. 2(b). As expected, the norm deviates nonlinearly with increasing σ and approaches zero at small σ. In order to confirm the quadratic deviation predicted by Eq. (11), the computed gradient and quadratic fit for the equivalent gradient on a simulated noise-free total SHG landscape is shown for reference. Good agreement with experimental values is obtained, and the deviation is quadratic and well-predicted by Eq. (11) for a spectral phase sampling domain size of $\lesssim 10\%$. The stabilization of the error curves in Fig. 2(b) at large σ is due to the influence of higher order terms in Eq. (6)

The convergence behavior for the statistical gradient will be dependent on the quantum system under study, the location of the sampling on the control landscape, and the operating conditions, including available shaper resolution and existing laser noise. However, the above measured convergence results qualitatively represent what may be reasonably expected of the statistical gradient on a quantum control landscape. Additionally, each individual system will have its own set of optimal sampling parameters, including a sufficient sample size N and σ value. Experience will aid in estimating N and σ, which should benefit from quantum control landscapes possessing inherently bounded (modest) slopes [10,11].

Having demonstrated the convergence properties of the method, an example of the recovered gradient form is presented. The spectral phase space for the filtered SHG landscape is sampled about a fixed phase function $[\phi(\Omega)=-3.7\times 10^3\ \mathrm{fs}^2\ \Omega^2]$ with a sample size of $N=50$ and a domain of $\sigma=2.5\%$ of 2π. The statistically recovered gradient $\frac{\delta S_f}{\delta\phi(\Omega)}$ is shown in Fig. 3 along with its theoretically predicted form [23]. It is readily observed that addition of the computed gradient to the original quadratic phase (scaled by a suitable step size) results in a spectral phase of decreased magnitude and, consequently, a higher filtered SHG signal.

The first-moment technique reproduces the gradient quite accurately, and it is especially useful to note that the gradient is estimated with a sample size smaller than the problem pixel dimension $D=64$. As discussed in the Appendix, the recovered gradient precision increases as $\sqrt{N}$, and enhanced precision may be required for demanding applications such as utilization of the gradient as a myopic spy to discover landscape structure.

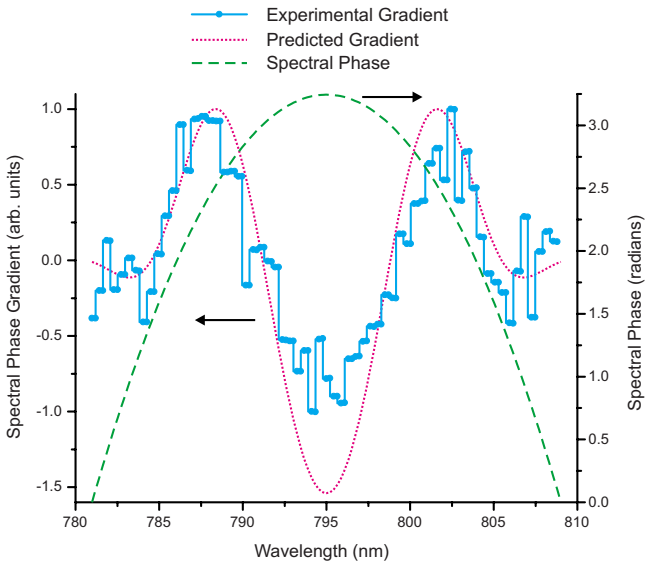

FIG. 3. (Color online) Recovered statistical gradient (solid line/blue line) for a point (i.e., reference spectral phase) on the filtered SHG landscape: $N=50$, $D=64$, and $\sigma=2.5\%$ of 2π. The theoretical gradient form (dotted line/red line) is shown for comparison along with the reference spectral phase (dashed line/green line).

V. CONTROL LANDSCAPE GRADIENT ASCENT ALGORITHM

A straightforward application of the measured gradient is implementation of a steepest-ascent algorithm to optimize quantum control problems. Climbing of the landscape is accomplished simply by proceeding in the steepest direction of the statistically determined gradient, i.e.,

$$\vec{x}^{(i+1)} = \vec{x}^{(i)} + \delta^{(i)} \times \vec{\nabla} J(\vec{x})^{(i)}, \tag{12}$$

where $\vec{x}^{(i)}$ is the landscape position at the ith iteration and $\delta^{(i)}$ is a scalar adjustable step size. For the presently considered systems, the control variables are represented by discrete values of the spectral phase, i.e., $x_i=\phi(\omega_i)$. At a given point on the quantum control landscape, the gradient is approximated with Eq. (7), where the variables x_i are drawn from independent Gaussian distributions.

Naturally, a given sampling size N gives sparse coverage as the dimensionality of the problem D increases; hence, there is an intrinsic trade-off between gradient accuracy and the efficiency of its measurement. However, landscape ascent is forgiving to gradient inaccuracy as any reasonable estimate of the gradient will likely result in taking an ascending step ahead. For the quantum control optimizations considered here, a sample size of $N=15$ is chosen, which provides a sparse sampling for the utilized pulse shaper dimensionality of $D=64$ pixels; however, even with this reduced sampling, rapid ascent is observed.

Generally, each problem will also have its own optimal sampling domain size σ, which must result in measurable

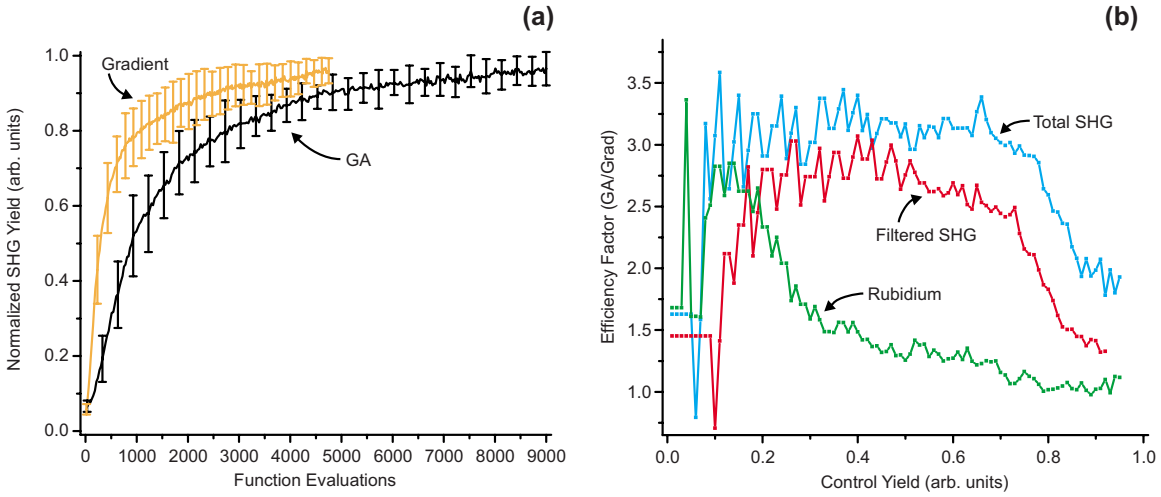

FIG. 4. (Color online) Experimental gradient and GA optimization of three model quantum systems: total SHG, filtered SHG, and atomic rubidium. Panel (a) displays the gradient (orange) and GA (black) optimization curves for total SHG. The two curves are averages of $n=14$ individual runs for each algorithm. Panel (b) displays the efficiency factor in terms of performed experiments (GA/grad) as a function of yield for each of the considered model quantum systems.

signal variations above the background noise fluctuations (see Appendix for general sampling domain considerations). For all problems treated here, the observed signal-to-noise ratio is quite high [24], and a sampling distribution standard deviation of $\sigma=0.05$ is chosen, which corresponds to variations in the spectral phase domain $\phi(\omega_i)$ of 0.314. The optimization is conducted in a spectral phase domain of $\phi(\omega_i) \in [0,2\pi]$ with these boundary conditions imposed by a phase wrapping operator. As mentioned earlier, no convergence difficulties are experienced since all sampling of the spectral phase is performed over an infinite domain and is only later constrained within this equivalent finite domain for experimental ease. Each variable $\phi(\omega_i)$ is allowed to vary continuously (within the resolution of the pulse shaper) and the achieved resolution is limited only by noise. Every optimization run, including those of the GA, begins at a random initialization point in the $D=64$ dimensional spectral phase space. The iteration dependent step size $\delta^{(i)}$ for the gradient is chosen by means of a simple one-dimensional bracketed line search using the golden ratio [21].

The performance of the steepest-ascent algorithm is compared to that of a traditional GA with bitstring representation of 6 bit resolution per pixel. The GA employs a fixed population size of 30 individuals, 2 crossover points, a mutation probability of $p_m=0.005$, and the selection mechanism is to keep the better half of the population while the single best individual is always retained (elitism). These parameters were collectively chosen to permit achievement of high-quality solutions with fast convergence.

A comparison between the standard GA and steepest-ascent efficiencies is made by recording the total number of necessary function evaluations for each algorithm, including those required for the bracketing routine. Here a function evaluation refers to the performance of an experiment with a new setting of the phase mask, and every function evaluation consisted of averaging over 500 single laser shots. In order to provide statistically meaningful results, a minimum of $n=12$ runs of each algorithm on a problem was completed, and all comparison-based inferences are drawn from the resultant collective behavior of the algorithms.

A representative comparison between the gradient and GA performances is shown in Fig. 4(a) for total SHG. As evident from the learning curves, the gradient achieves rapid ascent from the onset and performs at a rate approximately three times that of the standard GA [Fig. 4(b)]. The margin of its efficiency increase over that of the GA begins to decrease, however, in the yield range of ~70–80 %. This decrease is likely due to several coordinated factors, including a decrease in the actual magnitude of the gradient as the maximum on the landscape is approached; additionally, the signal variation ΔJ for a given sampling domain size σ decreases with the gradient magnitude, i.e., $\Delta J \simeq |\vec{\nabla} J(\vec{x})|\sigma$. Consequently, the precision of the measurement decreases as these variations approach the level of background noise fluctuations, and the number of samples N must be increased to maintain the same level of precision (see Appendix). Nonetheless, even with this observed decrease in convergence rate, the gradient algorithm converges to a maximum yield with $\sim 1/2$ of the function evaluations required by the GA.

The steepest-ascent performance may also be compared for the filtered SHG and rubidium model systems. For the former, an efficiency trend similar to total SHG is observed, namely, a saving of ~2.5 until the yield reaches ~75% at which point the final efficiency gain is ~1.5. In the case of rubidium, an initially modest efficiency gain of $\lesssim 2$ is observed until ~70%, at which point the algorithm performs on par with the GA.

The average yields obtained for each algorithm are shown in Table I. Although slower than the gradient, the GA is reliable and tends to obtain a slightly higher final yield. While the slowly varying landscape in the neighborhood of a maximum is an effective bottleneck for the gradient, the GA's stochastic nature is able to effectively maneuver this area in the presence of noise. Indeed, the majority of the gradient's gains are achieved in the first ~75% of the optimization. Thus, a natural course would be the marriage of

TABLE I. Performance analysis of the statistical steepest-ascent and GA learning algorithms on model quantum control problems. For each of the three systems, the average final yield and number of evaluations required to reach both 90% and 70% yields (the yield standard deviations are given in parentheses) are listed for both the gradient and GA. Yields for both total SHG and filtered SHG are defined in terms of the transform-limited pulse; yields for rubidium are normalized to a single optimal solution and thus are able to vary slightly about 1.0. These mean values are determined from an ensemble of n algorithmic runs. The experimental efficiency factors at 90% and 70% yields, defined by the ratio of GA to gradient evaluations, are shown for each system.

	Total-SHG ($n=14$)			Filtered-SHG ($n=12$)			Rubidium ($n=13$)		
Method	Avg. yield	Eval_{90}	Eval_{70}	Avg. yield	Eval_{90}	Eval_{70}	Avg. yield	Eval_{90}	Eval_{70}
GA	0.95(0.03)	5160	1740	0.96(0.03)	4620	1530	1.05(0.08)	3510	1530
Gradient	0.95(0.04)	2599	577	0.92(0.09)	3269	626	0.95(0.09)	3440	1323
Efficiency		1.99	3.02		1.41	2.44		1.02	1.16

deterministic and stochastic algorithms; the gradient could be utilized while the landscape is rapidly varying and large yield gains are possible, while a stochastic algorithm could be effective near the minimum and maximum of the landscape where the signal variations approach the noise limit. The gradient algorithm may also prove useful for optimization of short-lived systems (e.g., fragile biological systems or systems that bleach over time) when large gains are required in short time intervals. The observed algorithmic acceleration could also permit an increase in control dimension D for problems in which system degradation precludes the long optimizations required of a high-dimensional GA.

The intent of this work was not the development of a polished gradient algorithm for quantum optimal control but rather to present a proof of principle that the gradient can be measured and utilized. Thus, the chosen algorithmic parameters are not necessarily optimal, and numerous future improvements may be made. For example, incorporation of a state-of-the-art adaptive step-size mechanism specifically designed for handling noise [25] may significantly increase gradient performance. Moreover, one could use this methodology for implementation of a conjugate gradient algorithm in an attempt to attain further acceleration.

VI. SUMMARY AND OUTLOOK

As attention is increasingly turning toward the underlying fundamental quantum control landscapes [22,26,27] dictating experiment performance, it is necessary to develop reliable exploratory tools able to discern and navigate landscape structure in search of level sets [13], robust solutions [28], Pareto fronts [16,29], etc. The gradient flow, perhaps the most natural measure of landscape structure, provides insight into curvilinear control variable coordinate transforms, the source of constraints producing suboptimal traps, robustness to noise, and mechanistic insight. As such, the gradient is a myopic landscape spy, capable of reporting a number of experimental control properties desirable for analysis and comprehension of the optimization process.

To illustrate one such potential use of the gradient flow, a steepest-ascent algorithm for optimization of laboratory quantum control problems was implemented. Despite the fact that deterministic algorithms typically perform poorly in the presence of noise, evaluation of the gradient here is based on a robust stochastic technique; the gradient is measured by computing a weighted first moment from a local sampling of the landscape. Accordingly, the steepest-ascent algorithm displayed acceleration up to ~ 3 over that of the GA for yields up to $\sim 70-80$ % of the maximum. This reliable increase in convergence speed is a promising tool for future laboratory experiments, and extraction of underlying landscape trajectories may provide information about the particular optimization mechanisms.

The introduction of deterministic search would also allow an examination of the role that noise plays on the repeatability of quantum control optimizations. For instance, the steepest-ascent algorithm may be repeatedly run from a fixed starting point on the landscape; divergence from a prior trajectory may provide valuable information on the immunity or susceptibility of a quantum system and experimental setup to noise and day-to-day conditional fluctuations.

The mounting successes of quantum control experiments have begun to be explained in terms of the topology and structure of the underlying quantum control landscape, over which optimization algorithms operate, and knowledge of the gradient opens the door to a multitude of trajectory excursions over these landscapes, including exploration of level sets [12] and Pareto fronts [16]. This paper introduces a practical, easily implementable statistical method for obtaining the gradient on a general quantum control landscape. The introduction of this promising tool should aid in the continued navigation of quantum control landscapes in the laboratory with a goal of uncovering the properties that have made them so amenable to experimental optimization.

ACKNOWLEDGMENTS

The authors thank Ofer M. Shir for helpful discussions and algorithmic assistance and acknowledge support from the DOE.

APPENDIX

A brief analysis of the relevant signal-to-noise ratio issues pertinent to evaluation of the statistical gradient is given below. This analysis is meant to be neither comprehensive nor rigorous, but rather to provide appropriate estimates of the

necessary quantities. As such, laser noise is included as an additive scalar to the pulse intensity.

1. Sampling limits and statistical gradient accuracy for *m*-photon process

For a m-photon process, the observed signal $J^{(m)}$ may be approximately written as $J^{(m)} \propto I^m$, where I is the laser temporal intensity. Laser fluctuations may be taken into account by writing the measured signal as $J^{(m)}=(I_0+\delta)^m$, where I_0 is the unperturbed intensity and the additive factor δ is used to describe the laser noise. For convenience, the fluctuations δ are drawn from a normal distribution with mean zero and standard deviation of σ_n, i.e., $\delta \sim N(0,\sigma_n^2)$. The observed signal may then be written to first order in the noise δ as $J^{(m)} \simeq I_0^{(m)}+m\delta I_0^{(m-1)}$; the standard deviation of this m-photon signal is given by $\sigma^{(m)}=m\sigma_n I_0^{(m-1)}$, where $\langle \delta^2 \rangle = \sigma_n^2$ has been used. Thus, the m-photon signal-to-noise ratio may be written as $S/N^{(m)}=[S/N^{(1)}]/m$, where $S/N^{(1)}$ is the one-photon signal to noise, i.e., $S/N^{(1)}=I_0/\sigma_n$.

a. Sampling lower bound

The signal variations $\Delta J^{(m)}$ intentionally induced by statistically sampling the variate x must therefore be greater than the m-photon noise variation, i.e., $\Delta J^{(m)} > \sigma^{(m)}$. However, a small statistical sampling of the variate x affects the intensity in a linear fashion, i.e., $I(x) \simeq I_0+\frac{dI}{dx}x$, and is therefore considered to propagate through Schrödinger's equation in a similar manner as the laser noise δ. Essentially, the variate sampling can be viewed as a controlled introduction of "noise" and is accordingly considered as such. Consequently, the lower bound for the intentional sampling domain is the level of existing laser noise, i.e., $\sigma \gtrsim \sigma_n$.

b. Sampling upper bound

An upper bound for the variate sampling is established by requiring that the first-moment term dominates the Taylor expansion of Eq. (5). Consequently, it is required that

$$\mu_2 \frac{dJ^{(m)}(x_0)}{dx} \gtrsim \frac{\mu_4}{6}\frac{d^3J^{(m)}(x_0)}{dx^3}. \tag{A1}$$

Use of the m-photon model described above along with the assumption of a Gaussian sampling distribution (i.e., $\mu_2=\sigma^2$ and $\mu_4=3\sigma^4$) reduces this expression to

$$\sigma^2 \lesssim \frac{2I_0^2}{(m-1)(m-2)}. \tag{A2}$$

If the sampling domain σ is expressed as a fraction α of the laser noise σ_n, i.e., $\sigma=\alpha\sigma_n$, this upper bound may be written as

$$\alpha \lesssim \frac{\sqrt{2}S/N^{(1)}}{\sqrt{(m-1)(m-2)}} \simeq \frac{\sqrt{2}S/N^{(1)}}{m} = \sqrt{2}S/N^{(m)}. \tag{A3}$$

Thus, in order to maximize the accuracy of the first-moment method for estimating the gradient of a m-photon model, the fractional sampling domain α is constrained by the following bounds:

$$1 \lesssim \alpha \lesssim \sqrt{2}S/N^{(m)}. \tag{A4}$$

Hence, a high m-photon signal-to-noise ratio guarantees that the gradient is efficiently approximated with the statistical first-moment method and will not be severely contaminated by the landscape third derivatives.

c. Statistical gradient error

A lower bound for the fractional error intrinsic to the statistical gradient of a quantum control process may be derived from Eq. (6) to have a form

$$\frac{\left|\frac{\langle xJ^{(m)}(x_0+x)\rangle}{\sigma_n^2}-\frac{dJ^{(m)}(x_0)}{dx}\right|}{\frac{dJ^{(m)}(x_0)}{dx}} \gtrsim \frac{\sigma_n^2}{2}\frac{\frac{d^3J^{(m)}(x_0)}{dx^3}}{\frac{dJ^{(m)}(x)}{dx}}. \tag{A5}$$

With the m-photon model described above, this lower bound simplifies to

$$\frac{\left|\frac{\langle xJ^{(m)}(x_0+x)\rangle}{\sigma_n^2}-\frac{dJ^{(m)}(x_0)}{dx}\right|}{\frac{dJ^{(m)}(x_0)}{dx}} \gtrsim \frac{\sigma_n^2}{2I_0^2}(m-1)(m-2) \simeq \frac{m^2\sigma_n^2}{2I_0^2} = \frac{1}{2}\left(\frac{1}{S/N^{(m)}}\right)^2. \tag{A6}$$

Thus, the gradient inaccuracy increases as m^2 for a fixed sampling domain σ. The error intrinsic to the statistical gradient is limited by the signal-to-noise ratio at a particular location on the quantum control landscape.

It is worth noting that a more general model of the form $J^{(m)}=f(I^{(m)})$, where f involves an unknown functional dependence, yields similar results and reduces to the above form when $f(I^{(m)})=I^{(m)}$.

2. Statistical gradient precision of a *m*-photon process in the presence of noise

The estimator for the derivative $\frac{\partial J(x_0)}{\partial x}$ is given by $\frac{1}{\mu_2 N}\sum_n^N x_n J(x_0+x_n)$ and its mean value $\langle \frac{\partial J}{\partial x}\rangle$ is $\frac{\langle xJ(x_0+x)\rangle}{\mu_2}$ in the limit that $\sigma \to 0$. Hence, the estimator is unbiased [i.e., $\langle(\frac{\partial J(x_0)}{\partial x}-\langle\frac{\partial J(x_0)}{\partial x}\rangle)\rangle=0$], and the mean-squared error (E_{MS}) is given by

$$E_{\mathrm{MS}} = \left\langle \left(\frac{\partial J(x_0)}{\partial x} - \left\langle \frac{\partial J(x_0)}{\partial x} \right\rangle \right)^2 \right\rangle. \tag{A7}$$

By using the above mentioned form of the gradient estimator, the mean-squared error (E_{MS}) reduces to

$$E_{\mathrm{MS}} = \frac{\sigma_{xJ}^2}{{\mu_2}^2 N}, \tag{A8}$$

where σ_{xJ}^2 is the variance for the distribution of $xJ(x_0+x)$. Using the simple m-photon model discussed above, the measured signal in the presence of noise is written to first order as $J(x)=I_0^{(m)}+mI_0^{(m-1)}(x+\delta)$. The variance σ_{xJ}^2 may then be

expressed in terms of the sampling domain σ and experimental noise σ_n as

$$\sigma_{xJ}^2 \simeq \sigma^2 I_0^{(2m)}\left[1+\frac{2m^2\sigma^2}{I_0^2}\left(1+\frac{\sigma_n^2}{2\sigma^2}\right)\right]. \tag{A9}$$

The use of this variance in Eq. (A8) leads to a root mean-squared error of

$$\sqrt{E_{\mathrm{MS}}} \simeq \frac{I_0^{(m)}}{\sigma\sqrt{N}}\sqrt{1+\frac{2m^2\sigma^2}{I_0^2}\left(1+\frac{\sigma_n^2}{2\sigma^2}\right)}. \tag{A10}$$

It is convenient at this point to write a relative root mean-squared error to describe the gradient precision,

$$\frac{\sqrt{E_{\mathrm{MS}}}}{\langle J(x)\rangle} \simeq \frac{1}{\sigma\sqrt{N}}\sqrt{1+\frac{2m^2\sigma^2}{I_0^2}\left(1+\frac{\sigma_n^2}{2\sigma^2}\right)}. \tag{A11}$$

As expected and demonstrated in Fig. 2, the statistical gradient converges as $1/\sqrt{N}$. Interestingly, when the ratio $\frac{\sigma}{I_0}$ is below a certain threshold, the statistical gradient precision (i.e., number of requisite experiments) is independent of photon order m. This threshold is roughly established by requiring that the term $\sqrt{1+\frac{2m^2\sigma^2}{I_0^2}(1+\frac{\sigma_n^2}{2\sigma^2})}$ not deviate from a value of 1.0 by more than 10%, i.e.,

$$\frac{2m^2\sigma^2}{I_0^2}\left(1+\frac{\sigma_n^2}{2\sigma^2}\right) \lesssim 0.2. \tag{A12}$$

If the sampling domain standard deviation σ is again written as $\sigma=\alpha\sigma_n$, this expression simplifies to

$$S/N^{(1)} = \frac{I_0}{\sigma_n} \gtrsim m\sqrt{10\alpha^2+5}. \tag{A13}$$

For a reasonable choice of $\alpha=2$, the demand on single photon signal to noise becomes $S/N^{(1)} \gtrsim 6.7m$, which is certainly realizable for even high photon numbers m [24]. In this circumstance, the statistical gradient precision is not limited by the photon order m but only by the number of performed experiments N.

[1] S. Backus, C. G. Durfee III, M. M. Murnane, and H. C. Kapteyn, Rev. Sci. Instrum. **69**, 1207 (1998).
[2] A. Weiner, Rev. Sci. Instrum. **71**, 1929 (2000).
[3] T. Bäck, *Evolutionary Algorithms in Theory and Practice* (Oxford University Press, New York, 1996).
[4] R. S. Judson and H. Rabitz, Phys. Rev. Lett. **68**, 1500 (1992).
[5] J. Kunde, B. Baumann, S. Arlt, F. Morier-Genoud, U. Siegner, and U. Keller, Appl. Phys. Lett. **77**, 924 (2000).
[6] P. Nuernberger, G. Vogt, T. Brixner, and G. Gerber, Phys. Chem. Chem. Phys. **9**, 2470 (2007).
[7] A. Assion, T. Baumert, M. Bergt, T. Brixner, B. Kiefer, V. Seyfried, M. Strehle, and G. Gerber, Science **282**, 919 (1998).
[8] H. Rabitz, M. Hsieh, and C. Rosenthal, Science **303**, 1998 (2004).
[9] H. Rabitz, M. Hsieh, and C. Rosenthal, J. Chem. Phys. **124**, 204107 (2006).
[10] T.-S. Ho and H. Rabitz, J. Photochem. Photobiol., A **180**, 226 (2006).
[11] H. Rabitz, T. S. Ho, M. Hsieh, R. Kosut, and M. Demiralp, Phys. Rev. A **74**, 012721 (2006).
[12] A. Rothman, T.-S. Ho, and H. Rabitz, Phys. Rev. A **73**, 053401 (2006).
[13] J. Roslund, M. Roth, and H. Rabitz, Phys. Rev. A **74**, 043414 (2006).
[14] A. Rothman, T.-S. Ho, and H. Rabitz, Phys. Rev. A **72**, 023416 (2005).
[15] V. Beltrani, P. Ghosh, and H. Rabitz J. Chem. Phys. **130**, 164112 (2009).
[16] R. Chakrabarti, R. Wu, and H. Rabitz, Phys. Rev. A **77**, 063425 (2008).
[17] T.-S. Ho (private communication).
[18] $f(\sigma)=\sigma^3[(1+2\sigma^2)(1.5+\frac{1}{\sigma^2}+\frac{1}{2+4\sigma^2})]^{3/2}$; $\lim_{\sigma\to 0} f(\sigma)=1$.
[19] D. Meshulach and Y. Silberberg, Phys. Rev. A **60**, 1287 (1999).
[20] N. Dudovich, B. Dayan, S. M. Gallagher Faeder, and Y. Silberberg, Phys. Rev. Lett. **86**, 47 (2001).
[21] W. H. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes* (Cambridge University Press, New York, 1992).
[22] J. Roslund and H. Rabitz, Phys. Rev. A (to be published).
[23] The spectral phase gradient for filtered SHG S_f at twice the carrier frequency, i.e., $\Omega_{fg}=\omega_{fg}-2\omega_0=0$ is shown to be $\frac{\delta S_f}{\delta\phi(\Omega)}=-4A(\Omega)A(-\Omega)R_{fg}\sin[\phi(\Omega)+\phi(-\Omega)-\alpha_{fg}]$, where $A(\Omega)$ is the fundamental spectral amplitude and R_{fg} and α_{fg} are the spectral amplitude and phase of the doubled field, respectively [22].
[24] M. Roth, J. Roslund, and H. Rabitz, Rev. Sci. Instrum. **77**, 083107 (2006).
[25] N. Hansen and A. Ostermeier, in *Proceedings of the 1996 IEEE International Conference on Evolutionary Computation* (IEEE, Piscataway, 1996), pp. 312–317.
[26] M. Wollenhaupt, A. Präkelt, C. Sarpe-Tudoran, D. Liese, and T. Baumert, J. Mod. Opt. **52**, 2187 (2005).
[27] T. Bayer, M. Wollenhaupt, and T. Baumert, J. Phys. B **41**, 074007 (2008).
[28] A. F. Bartelt, M. Roth, M. Mehendale, and H. Rabitz, Phys. Rev. A **71**, 063806 (2005).
[29] L. Bonacina, J. Extermann, A. Rondi, V. Boutou, and J. P. Wolf, Phys. Rev. A **76**, 023408 (2007).