Optics Communications 284 (2011) 3748–3758

Contents lists available at ScienceDirect

Optics Communications

journal homepage: www.elsevier.com/locate/optcom

Investigation of adaptive laser pulse shaping by direct estimation of group delay profile

Di Yang, Daan P. Sprünken, Alexander C.W. van Rhijn, Janne Savolainen, Ting Lee Chen, Herman L. Offerhaus, Jennifer L. Herek, Aliakbar Jafarpour *

Optical Sciences group, MESA+ Institute for Nanotechnology, University of Twente, Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 27 October 2010
Received in revised form 24 February 2011
Accepted 27 February 2011
Available online 14 March 2011

Keywords:
Ultrafast optics
Pulse shaping
Phase
Delay
Evolutionary algorithm

ABSTRACT

We investigate the benefits of estimating the group delay spectrum (as opposed to the phase spectrum) on the performance of an adaptive laser pulse shaping experiment. We use both numerical and experimental approaches, and consider different figures of merit such as efficiency, noise robustness, convergence speed, repeatability, fitness sensitivity, and parameter insensitivity. While primarily focused on the process of broadband second harmonic generation, we also study a resonant third order nonlinear process. Our results show an overall superiority of delay-based formulation in the cases under study (compared to a conventional phase-based formulation) for adaptive laser pulse shaping.

1. Introduction

Tailoring laser pulse properties, referred to as pulse shaping, is increasingly used in ultrafast spectroscopy to induce classical or quantum interferences and to control the excitation–relaxation processes in atomic and molecular systems [1–3]. Given the insufficient knowledge about the complex underlying dynamics and/or the control mechanism in many practical optimizations, pulse shaping is commonly implemented using an adaptive scheme [4–8]. Despite successful demonstration of the phase-sensitive nature of many light-matter interactions using adaptive schemes, efficient implementations of optimization algorithms in such experiments and the extraction of (even qualitative) insight from the "optimal" phase profiles are still key challenges [9–12].

A common depiction of pulse shaping describes a pulse with the magnitudes and the relative arrival times of the constituent wavelengths. These two features are usually quantified by the amplitude and the *phase* of the pulse in its frequency domain formulation. Here we show that directly modeling the *group delay*, rather than coding it in phase, can benefit adaptive pulse shaping at both fundamental and practical levels.

Some benefits of delay-based pulse shaping are as follows. First, delay is a physical observable with intuitive meaning; while phase is an abstract and congruent quantity (different phase values may be considered equivalent modulo 2π). Second, delay-based modeling can be used for a one-to-one mapping of delay to phase over phase spans exceeding 2π, and an efficient use of phase masks providing such spans. Third, a direct estimation of delay eliminates the challenges of phase unwrapping and numerical artifacts of differentiation, encountered in obtaining the delay profile from an estimated phase profile. The reader interested in a more fundamental treatment of this subject is referred to [13] for contrasting phase and delay in coherent control, and to [14] for an investigation of nontrivial ambiguities of relative phase and delay associated with isolated spectral components.

Here, we investigate the performance of delay-based *adaptive optimizations* in terms of efficiency, noise robustness, convergence speed, repeatability, fitness sensitivity, and parameter insensitivity in this contribution. We conduct a comprehensive study of "differential basis" reported in some earlier experiments in adaptive laser pulse shaping (1D) [15], which is comparable to a similar approach in image processing (2D) for the estimation of the unwrapped phase profile [16]. Our numerical and experimental results show that in most cases and under similar conditions, a delay-based optimization outperforms a phase-based optimization.

2. Basic concepts

2.1. Adaptive laser pulse shaping with evolutionary algorithms

Numerical optimizations, such as evolutionary algorithms, are based on finding patterns in numerical data (top–down approach), rather than

* Corresponding author. Tel.: +31 53 489 3290; fax: +31 53 489 3511.
E-mail address: jafarpoura@gmail.com (A. Jafarpour).

0030-4018/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.optcom.2011.02.075

an analytical formulation of the problem or the optimization strategy (bottom–up approach).

Evolutionary algorithms explore an optimization space by trying a set (here referred to as *population*) of candidate solutions. The relative success of optimization for each candidate solution is quantified by a figure of merit (*optimization fitness*). By knowing (simulating or measuring) the fitness values of different candidate solutions and applying some stochastic statistical operations in each iteration (*generation*), a new set of candidate solutions is generated. Quantifying the minimum, the average, and the maximum fitness values in each generation gives an impression of the relative spread of the population in the search space, or equivalently the richness (flat vs. structured shape) of the optimization space at points corresponding to the current population. Furthermore, tracking the average (or the minimum/maximum) fitness in consecutive generations (*learning curve*) gives qualitative and quantitative insights regarding the optimization. One main aspect of an optimization is its vulnerability to being trapped in local optima, instead of reaching the global optimum. This issue will be addressed in Sections 3.3, 3.4, and 4.3.2. The stochastic nature of these optimizations implies that even in the absence of noise and with exactly similar parameters, the outcomes of several runs of the same program can be different, as different realizations of random number sequences are employed in calculations.

Here, we use an evolutionary algorithm with promising results in adaptive laser pulse shaping, namely the covariance matrix adaptation evolution strategy (CMA-ES) [17]. Details of implementation of the code have been reported before [18].

Fig. 1(top) shows the block diagram of an adaptive laser pulse shaping system. The pulse shaper block adds a phase profile $\varphi_{shaper}(\omega)$ to the phase of the input laser pulse $\varphi_{in}(\omega)$. The shaped pulse with a net spectral phase of $\varphi_{net}(\omega)=\varphi_{in}(\omega)+\varphi_{shaper}(\omega)$ is then applied to the optical process. The optimization algorithm uses feedback from the experiment to generate a new set of parameters, which determine φ_{shaper} and φ_{net} via a given parameterization. The optical process in our study is broadband second harmonic generation, which is optimized when $\varphi_{net}(\omega)=0$. We will also address a different (resonant) optical process in Section 5.

2.2. Optical process: second harmonic generation

We consider collinear type I phase-matched second harmonic generation (SHG) of a laser pulse in the non-depleted pump regime, as the optical process shown in Fig. 1. The pulse is incident normally on a thin uniaxial crystal (phase-matching bandwidth considerably larger than the pulse bandwidth), with instantaneous χ^2 nonlinearity and ideal anti-reflection coatings on both sides. Under these conditions and with the slowly-varying envelope approximation, the nonlinear component of the output electric field (referred to as the SHG signal) is proportional to the square of the input electric field in the time domain. Equivalently, the complex envelope of the SHG electric field is the autoconvolution of the fundamental electric field envelope in the frequency domain [19].

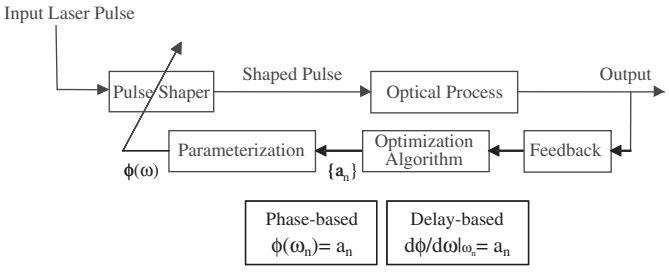

Fig. 1. (Top) Block diagram of an adaptive pulse shaping experiment, and (bottom) two different parameterization schemes. Algorithm outputs $\{a_n\}$ may first be shifted and/or scaled by the parameterization block (before being assigned to the phase variable or its derivative).

By applying different phase profiles to the input laser pulse, different profiles and total energies of the SHG signal can be obtained. We define the optimization fitness to be the ratio of the energy of the SHG field associated with a given phase function, normalized to that of a zero-phase pulse. Maximizing this fitness function, referred to as SHG optimization, is equivalent to compensation for the phase of the input laser pulse and achieving a transform-limited shaped pulse in Fig. 1 [19]. SHG optimization is commonly used in pulse shaping spectroscopy as a routine preliminary step to achieve and/or verify a flat-phase pulse at the sample position before shaping the pulse for the main experiment. It is also often used as an independent test case to evaluate an adaptive pulse shaping system.

2.3. Significance of parameterization

An important aspect of the CMA-ES algorithm is incorporating only a few internal parameters, with almost similar optimal values (or ranges) for different optimization problems [20]. As such, once an optimal implementation of the algorithm is found, one can change the parameterization block in Fig. 1(top) for different experiments, without changing the algorithm parameters. In this way, an efficient implementation of a given optimization is considerably disentangled from an efficient implementation of the algorithm itself. It will also enable one to use an efficient parameterization (of a given process) with alternative algorithms [21] with little or no modification.

Fig. 1(bottom) shows two different implementations of the parameterization block. The common phase-based parameterization uses the algorithm output parameters $\{a_n\}$ as samples of the phase profile to be programmed on the pulse shaper (φ_{shaper}). The delay-based parameterization, the subject of this study, uses the algorithm output parameters $\{a_n\}$ as samples of the derivative of φ_{shaper}. In both parameterizations, intermediate values of the sampled profile are obtained by interpolation (in our study by spline interpolation), and algorithm outputs may also be shifted and/or scaled (see Section 3.2).

This contribution is primarily focused on the relative success of the two (delay- and phase-based) parameterizations under similar conditions, and not a comparison of different conditions (such as linear vs. spline interpolation) or alternative (basis-set [11] or model-based [22]) parameterizations.

2.4. Significance of range and number of parameters

The range of algorithm parameters should be large enough to ensure an estimated profile with large enough phase (or delay) span. On the other hand, an inappropriately-large range of parameters can slow down the algorithm. The pixelated nature [23] and the finite temporal range of a pulse shaper and also the finite dephasing time of an excited state [24] put an upper bound (as small as sub-ps) on the parameter span, too. The number of parameters should also be sufficiently high to estimate the target profile with reasonable resolution, and sufficiently small not to slow down the algorithm ("curse of dimensionality" [20]). As such, the performance of adaptive optimizations shown in Fig. 1 is best assessed by a simultaneous study of the effects of the range and the number of algorithm output parameters. Similar limitations exist in optimizations using basis set parameterization (see Fig. 6 in [11]).

In many adaptive pulse shaping studies, a relatively large number of parameters (compared to the order of the target polynomial, for instance) are used to achieve high fitness values. Even if such high fitness values are achieved in practical spectroscopic problems in the presence of noise, there are two fundamental issues associated with such formulations. First, the slow down of the algorithm by the increased number of parameters is a serious issue, while facing laser instabilities and sample degradations. Second, the imperfect

estimation of phase wraps in the estimated wrapped phase profile makes it difficult to unwrap the phase for the calculation of the group delay or even a qualitative understanding.

Here, we limit our study to a relatively small number of parameters (up to 30) and investigate the relative performance of delay- and phase-based optimizations under similar conditions in this regime. As in a pioneer study of adaptive pulse shaping [25], we perform our comparative studies by limiting the number of generations. The number of function evaluations (used as a common cost index) is simply proportional to the number of generations, since the population size is constant (40) in all our optimizations.

2.5. Phase vs. delay

While the group delay can be observed directly (in the time domain) at microwave and lower frequencies, it is common to estimate the group delay indirectly (in the frequency domain) by differentiating the *unwrapped* phase profile at optical frequencies. The group delay can also be extracted directly from a spectrogram (time-frequency domain) at optical frequencies [26].

Our study is limited to many-cycle laser pulses, for which the constant term of the spectral phase can be reasonably discarded. Hence, the constant of integration, transforming the group delay profile to the phase profile, can be chosen arbitrarily. Note that in many cases (and possibly also in adaptive optimizations), an efficient modeling of pulse distortions in the single-cycle pulse regime [27] depends on explicit consideration of the constant phase term.

We employ the same hardware and calibration parameters (in experiments) and also the same code (in simulations) to program spectral phase profiles on our phase mask, in the case of both phase- and delay-based optimizations. The difference between these two optimizations is simply the existence of an integrator in the feedback path (or lack there of), or equivalently in the quantity *directly* estimated by the algorithm.

3. Simulation results and discussions

3.1. Simulation parameters

The simulation domain is scaled and discretized similar to the shaping coordinate (pulse shaper) with 640 points and a wavelength calibration of 0.155 nm/pixel. It is then re-sampled to have a uniform sampling rate ($d\omega = 0.654$ rad/ps) required for Discrete Fourier Transform (DFT) calculations, and also centered at zero to correspond to the spectrum of the envelope field. The input spectrum is a Gaussian with a full-width at half-maximum (FWHM) bandwidth of 35 nm centered at 670 nm, and has a phase profile $\varphi_{in}(\omega)$ with quadratic ($\varphi_2 = 200$ fs^2), cubic ($\varphi_3 = -4000$ fs^3), and quartic ($\varphi_4 = -10{,}000$ fs^4) phase terms. It features a maximum and a minimum point, positive and negative curvatures, and a span exceeding 2π. We refer to this phase profile as the *target* phase profile in the rest of this article.

The CMA-ES algorithm settings (except the size of the output vector; i.e., the number of a_n parameters in Fig. 1) are similar to those used in our previous studies [18]; i.e., a population size of 40, a parent number of 20, an initial step size of 10%, and parameter values limited to the interval [0,1]. Given the stochastic nature of evolutionary algorithms, any simulation program is executed 20 times, and the results are averaged to find the expected value of the solution and to verify the repeatability of optimizations.

3.2. Effect of range and number of parameters

Here, we try to quantify and to contrast the sensitivities mentioned in Section 2.4 in the cases of phase- and delay-based optimizations. We consider the number of parameters from the set {3,4,5,8,10,12,15,20,30} to model the 4-th order polynomial describing the target phase. The parameterization block in Fig. 1 converts the algorithm output parameters a_n to phase samples as $\varphi(\omega_n) = 2\varphi_M(a_n - 0.5)$, or to delay samples as $\tau(\omega_n) = 2\tau_M(a_n - 0.5)$. In order to have a fair comparison between the ranges of parameters for delay- and phase-based optimizations, we take two points into consideration. First, the profiles estimated directly in both cases (delay and phase) are related via the linear operator of differentiation. So, the two parameters τ_M and φ_M should vary over similar spans; i.e., $\varphi_{M,Max}/\varphi_{M,Min} = \tau_{M,Max}/\tau_{M,Min}$. Fig. 2 shows such a common span of 4 octaves ($\varphi_{M,Max}/\varphi_{M,Min} = \tau_{M,Max}/\tau_{M,Min} = 2^4 = 16$) for fitness values after 60 generations. Second, the minimum values of τ_M and φ_M should be "equivalent". We match the values of "small" τ_M and φ_M such that they have the same maximum fitness (obtained by varying the number of parameters). This common maximum fitness is considerably smaller than unity, and is arbitrarily chosen to be ~0.87. The corresponding values of parameters are $\tau_M = 15$ fs and $\varphi_M = \pi/4$.

Many intermediate values of τ_M and φ_M (corresponding to the two panels of Fig. 2) have also been considered, and corresponding optimizations have been simulated. They simply show a gradual shift from the red curves towards the blue and then towards the black curves. As such, these intermediate curves have not been plotted for the sake of clarity.

Comparing the two panels of Fig. 2 shows significant insensitivity of delay-based optimization to the number and the range of parameters, compared to a phase-based optimization. The final fitness (after 60 generations) in delay-based optimizations with $\tau_M = 50$ fs has an almost unity average and very small standard deviation, almost independent of the number of parameters. High fitness is achieved by phase-based optimization only with 4 and 5 parameters. The final fitness then drops significantly (compared to the delay-based case) depending on the number of parameters. Comparing the largest values of parameter in both cases (black curves on each panel) shows significant sensitivity and also less repeatability (large standard deviation) of phase-based optimization (right panel). Note that a given value of standard deviation (σ_F) is associated with a larger value of normalized deviation ($\sigma_{F,norm} = \sigma_F/F_{mean}$) at smaller values of the mean value (F_{mean}).

3.3. Disentangling the speed factor

The reduced fitness for large numbers of parameters in Fig. 2 can be the result of more structured landscapes and being trapped in local optima, or simply a signature of the curse of dimensionality (slow convergence). In order to disentangle these two mechanisms, we consider the largest number of parameters (30) and the largest parameter values ($\tau_M = 250$ fs and $\varphi_M = 4\pi$) from the simulation parameters shown in Fig. 2. We then repeat the simulations by increasing the number of generations to 200, 400, and 800. The distribution of the final fitness values for all these numbers of generations is shown in Fig. 3(left). The complete distribution of the final fitness values (20 runs of a given simulation program) after 800 generations is also shown in Fig. 3(right).

Fig. 3(left) shows the better performance of delay-based optimization in the sense of higher mean values and smaller (and decreasing) standard deviations of the final fitness. The increased value of the mean fitness by going from 60 to 200 generations (for both parameterizations) shows a partial contribution of the curse of dimensionality. On the other hand, the existence of a considerable standard deviation (even after 800 generations) implies structured landscapes (occasional vulnerability to local optima), especially in the case of phase-based parameterization.

Note that slower optimizations with large parameters can be because of not only the two aforementioned factors, but also their combined effect. Even if the algorithm manages to emerge out of a local optimum and reach the global optimum, this "detour" still takes

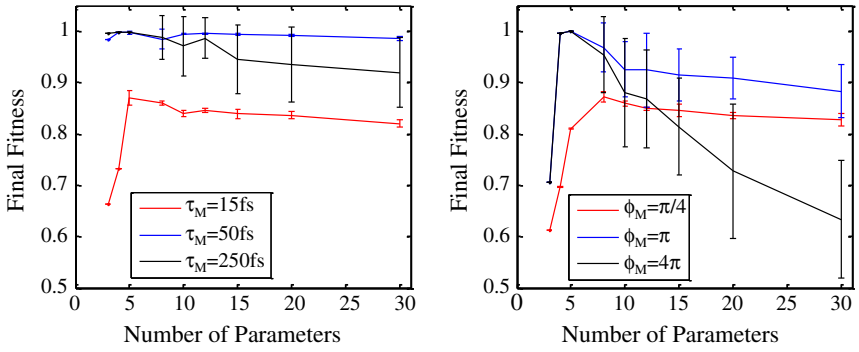

Fig. 2. The effect of the number and the range of algorithm output parameters on the final optimization fitness in the case of (left) delay-based and (right) phase-based optimizations under otherwise-similar conditions. The panels show the effect of small (red), intermediate (blue), and large (black) values of parameters over a span of 4 octaves. Simulation results for each case (pair of range-number of parameters) are represented by the average and the standard deviation of the distribution of final fitness values (obtained from 20 optimizations). The solid lines connecting the data points are guides to the eye.

the algorithm a few generations to concentrate (the population of solutions) around a local optimum and then leave it. This behavior may also partially account for a counter-intuitive decrease of the mean final fitness vs. the number of generations, as in Fig. 3(left) by going from 400 to 800 generations. Such a decrease can also originate from the large standard deviation of fitness values and the dependence of the mean final fitness on the specific ensemble used for averaging. Note that the results shown in Fig. 3(left) for different numbers of generations are all obtained by different optimizations, and without recording intermediate results.

An interesting feature in Fig. 3(left) and also both panels of Fig. 2 is the extension of error bars beyond the maximum fitness of unity. Note that standard deviation is an intuitive measure of spread for a random variable with a normal (or a similar bell-type) distribution profile. Error bars extended beyond unity imply an asymmetric distribution of final fitness values, with a major contribution to variance from a "considerably small" value (the majority of the distribution being greater than the mean value). Fig. 3(right) shows such a distribution for delay-based parameterization. The mean and the standard deviation of this distribution form the point corresponding to 800 generations in Fig. 3(left).

3.4. Fitness distribution and implication of local optima

As mentioned in Section 2.1, the stochastic nature of evolutionary optimizations can result in different fitness values for different runs of a given optimization (fixed parameters, yet different realizations of random variables). Geometrically, such different fitness values are associated with different optimization trajectories in the optimization space that end up in different local/global optima. Here, we take a closer look at delay- and phase-based optimizations from this perspective. In order to get better insight into the meaning and significance of fitness distributions, we run the simulation program corresponding to delay-based optimization with $\tau_M = 500$ fs and 15 parameters several times. We notice three patterns in the results. These three patterns are shown by plotting the associated learning curves (fitness vs. generation index) and the final estimation of the phase profile after 200 (as opposed to 60) generations. With a considerably larger number of generations and in the absence of noise, a distribution of the final fitness implies the existence of local optima and the probability of bypassing them during the search for the global optimum.

Fig. 4(top-left) shows the fastest learning towards the global optimum without a noticeable slow down. Fig. 4(top-middle) shows a slower learning towards the same maximum fitness with an obvious approach to a local optimum around the 50th generation. Finally, Fig. 4(top-right) shows an even slower learning towards a smaller final fitness value by emerging out of at least two noticeable local optima.

3.5. Noise robustness

Numerical differentiation can enhance numerical errors [28], while integration can have the opposite effect. Here, we investigate the potential noise suppression impact of the integration operator, required to obtain the spectral phase profile from the group delay profile, on our stochastic optimizations.

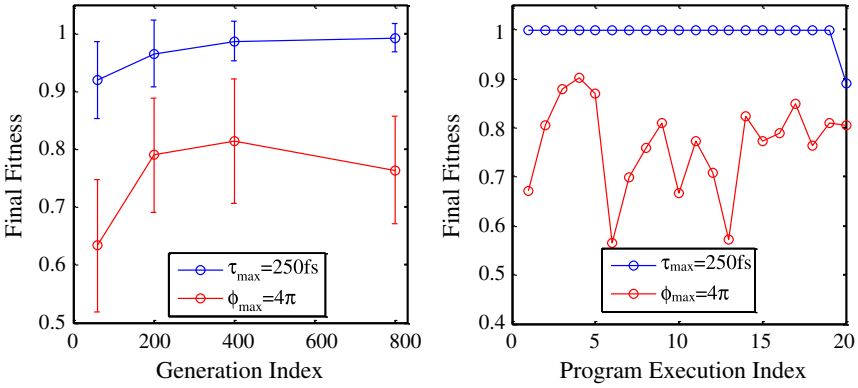

Fig. 3. (Left) The effect of longer optimization time on the achieved final fitness in delay- and phase-based optimizations with 30 parameters and the largest parameter values ($\tau_M = 250$ fs and $\varphi_M = 4\pi$, respectively), and (right) the distributions corresponding to the error bars with 800 generations shown on the left panel. These distributions show the final fitness values achieved by 20 runs of the same optimization problem in each case.

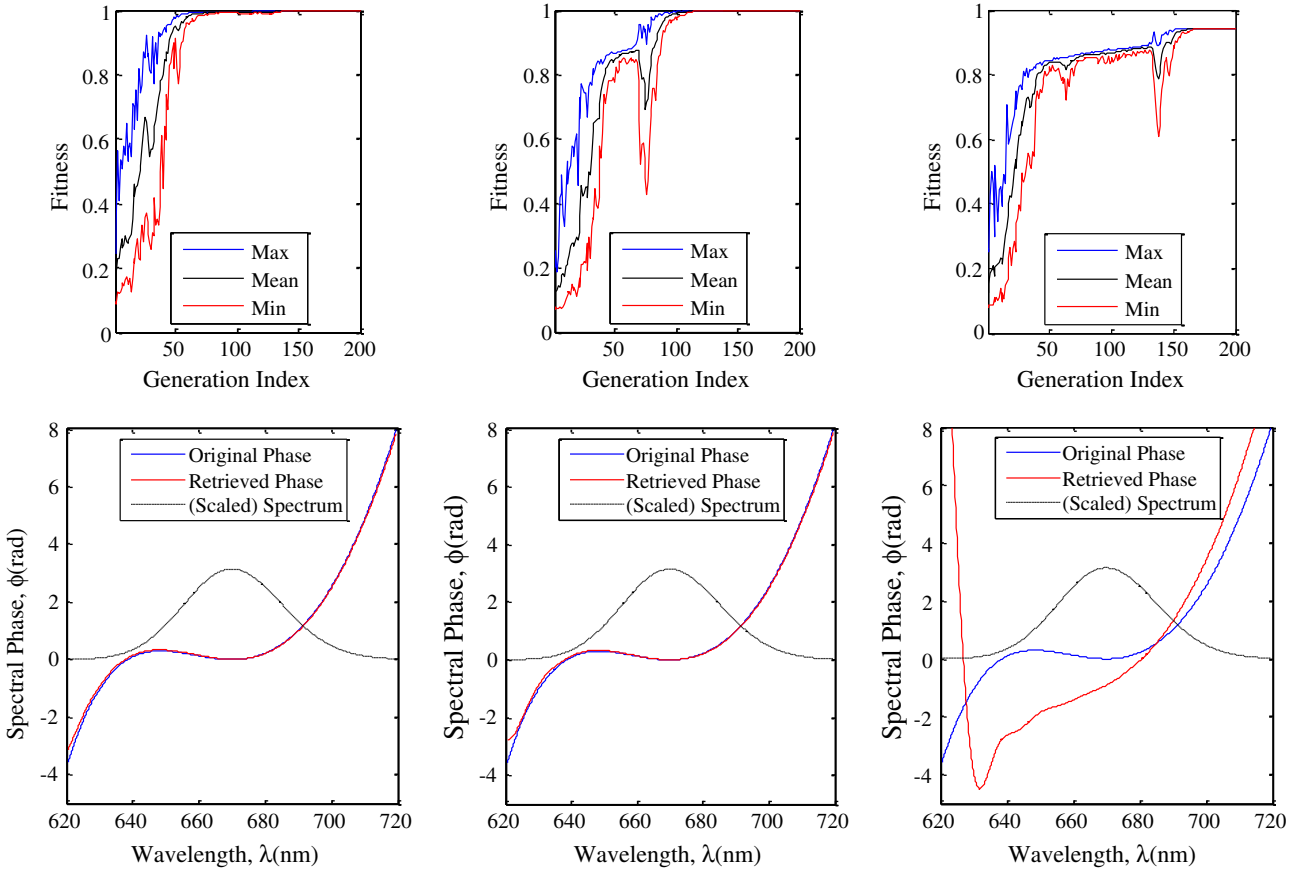

Fig. 4. Demonstration of (top) the learning curves and (bottom) the estimated phase profile after 200 generations in 3 runs of the same optimization program. All parameters are the same in the 3 cases, and the differences merely arise from a stochastic search (similar parameters, but different random number realizations) in a structured search space. The final estimated phase profiles have been shown along with the target phase profile and the pulse spectrum. The optimization program uses delay-based parameterization with 15 parameters and a parameter range of $\tau_M = 500$ fs.

We repeat the simulations described in Section 3.2 for some values of parameters by adding a phase noise [12] to the spectral phase of the laser pulse. The phase noise is the realization of a zero-mean and unity-variance random variable with normal (Gaussian) distribution, which is multiplied by a factor φ_N to introduce different noise levels. The numerical results of estimating the target phase with phase-based and delay-based optimizations at different levels of phase noise are shown in Fig. 5. As in previous simulations, each simulation program is executed 20 times, and the results are averaged. The noise signal is generated in the innermost loop of the program (each time with a different seed of the pseudo-random number generator) to ensure that no two realizations of the noise signal are the same (except statistically) [11].

Comparing the plots of the first row in Fig. 5 with corresponding plots of the second row suggests a superior performance of the delay-based optimization compared to the phase-based optimization (except for the second column). This superiority is both in terms of comparison with the simulated target phase (accuracy) and insensitivity to the noise level (precision).

The second plot on the first row (10 parameters, $\tau_M = 250$ fs) seems to have the worst performance among delay-based optimizations, and it is not easy to compare it with the corresponding phase-based optimization results just by visual inspection. As such, we also plot the mean and the standard deviation of optimizations with different parameterizations in a single graph for easy comparison. Fig. 6 shows that in all these cases, delay-based optimizations outperform phase-based optimizations in the sense of the mean value, the insensitivity to the number of parameters, and in most cases, also in the sense of smaller standard deviation.

Note that the fitness values corresponding to different noise levels cannot be readily compared with each other, as (the expected value of) the maximum fitness is a noise-dependent value, and not unity [12]. Also care must be taken in the interpretation or generalization of these results, in the case of other types or levels of noise.

4. Experimental results and discussions

4.1. Setup

In order to verify the simulation results in practice and in the presence of different sources of disturbance, several experiments with different parameterizations are conducted. The experimental setup has been described previously [5]. In brief, laser pulses centered at 670 nm with a FWHM bandwidth of 35 nm are generated by a non-collinear optical parametric amplifier (NOPA; Clark-MXR), which is pumped by an amplified femtosecond laser system (Clark CPA-2001; Clark-MXR). The pulse shaper uses a dual-stack liquid-crystal spatial light modulator (SLM-1280-VN-R; Cambridge Research & Instrumentation Inc.) placed in the Fourier plane of a 4-f ($f = 500$ mm) zero-dispersion compressor with two 1200 grooves/mm gratings and cylindrical mirrors, leading to the resolution of 0.155 nm/pixel. Only one mask of the SLM with 640 pixels is used for phase-only pulse shaping. A concave mirror ($f = 250$ mm) focuses the output of the pulse shaper, with a FWHM diameter of 2 mm and energy of 210 nJ/pulse, onto a β-barium borate (BBO) crystal with a thickness of 20 μm. The SHG signal is collimated and then coupled into a spectrograph (SP2150; Princeton Instruments) using a parabolic mirror with $f = 150$ mm. A 256-pixel diode array converts the optical output of

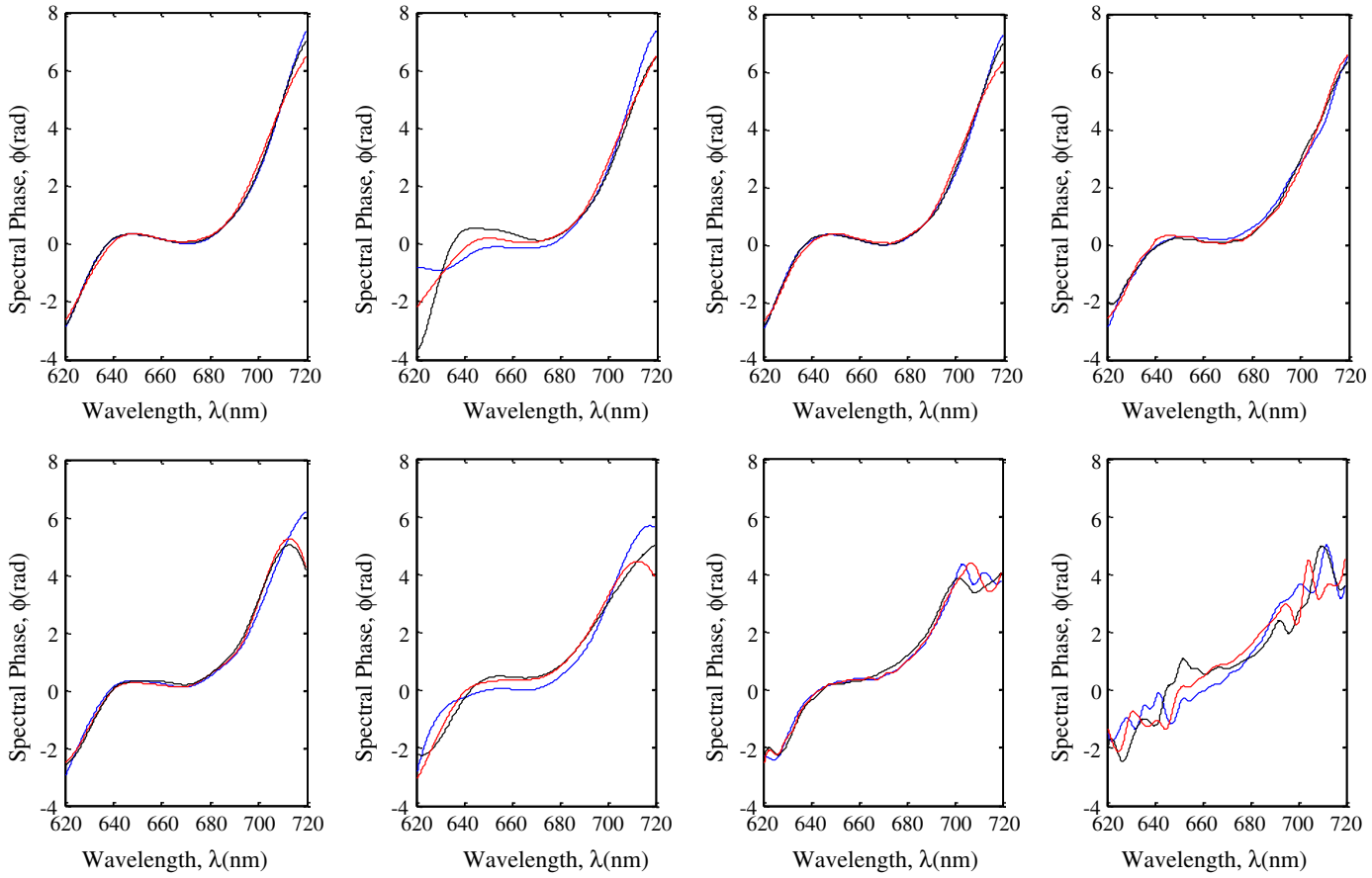

Fig. 5. The impact of phase noise on the estimation of the target phase with (top) delay- and (bottom) phase-based optimizations. The blue, the black, and the red curves correspond to phase noise amplitudes of $\varphi_N=0$, $\varphi_N=0.4$, and $\varphi_N=0.8$. The number and the maximum value of parameters corresponding to the top row and from left to right are ($n=10$, $\tau_M=50$ fs), ($n=10$, $\tau_M=250$ fs), ($n=30$, $\tau_M=50$ fs), and ($n=30$, $\tau_M=250$ fs). The number and the maximum value of parameters corresponding to the bottom row and from left to right are ($n=10$, $\varphi_M=\pi$), ($n=10$, $\varphi_M=4\pi$), ($n=30$, $\varphi_M=\pi$), ($n=30$, $\varphi_M=4\pi$), for phase-based optimizations. The target phase profile is similar to the one shown in Fig. 4, and is not shown here for the sake of clarity.

the spectrograph into electrical signals, which are then sampled at a rate of 1 kHz to have single-shot statistics. The fitness is calculated by integrating the weighted SHG spectrum. The weighting is simply dividing an SHG spectrum by a Gaussian profile (similar for all experiments) to enhance the contribution of the tails of the pulse to the total energy [11].

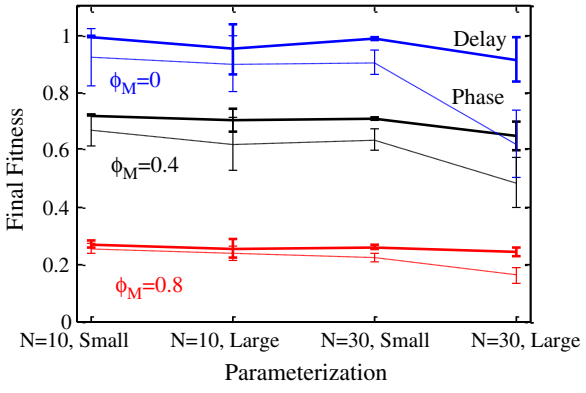

Fig. 6. The impact of phase noise on the final optimization fitness using phase- and delay-based approaches with four combinations of the number and the range of parameters in each case. The number of parameters is denoted by N, and the range of parameters is expressed as Small ($\tau_M=50$ fs or $\varphi_M=\pi$) or Large ($\tau_M=250$ fs or $\varphi_M=4\pi$). Different colors represent optimization results with different noise levels. The points connected by solid (dash) lines correspond to delay-based (phase-based) optimizations. The lines are guides to the eye, only.

4.2. Data validation

The fitness with zero added phase (reference fitness) is measured along with the fitness measurements for each generation of each optimization. The reference fitness corresponds to a pulse obtained by programming a flat phase on the pulse shaper, and not a shaped pulse with a net phase of zero. Ideally, the reference fitness values should be equal during an optimization. In practice, the plot of the reference fitness vs. generation is a horizontal line with some fluctuations. Assuming that such variations originate from slow and small variations of the spectrum (drift with no phase distortion), one can normalize the fitness values of each generation to the corresponding reference fitness to find comparable fitness enhancements of different generations. If the fluctuations of the reference fitness have large amplitudes or occur rapidly, it will become more difficult to disentangle the learning process of the algorithm and laser fluctuations. Fitness fluctuations can also originate from phase instabilities, even with a stable spectrum (amplitude). Normalizing the fitness values in such a case will be non-trivial, if not impossible.

An important practical issue in nonlinear spectroscopy with shaped pulses is the degradation of the spatial (M^2 parameter), temporal (shot-to-shot energy and phase stability and also drift), and the spatio-temporal (spatial chirp and pulse front tilt) characteristics of optical pulses along the nonlinear conversion pipeline; i.e., from the femtosecond oscillator to the regenerative amplifier, and then to the

parametric amplifier. In general, a successful "experimental result" obtained by use of a femtosecond oscillator can be more difficult to repeat with the same success level using a regenerative amplifier, and even more difficult with a parametric amplifier. We have chosen a parametric amplifier in this study to investigate the performance of the algorithm under the same condition that our molecular spectroscopic studies are performed.

The results of many optimization experiments are discarded, because of relatively large drifts (a standard deviation more than ~8% of the reference fitness, as opposed to <2% for the most stable reference fitness profiles). In all cases, single-shot measurements of the energy of the fundamental pulse show stabilities better than 1%. With an almost constant energy of the fundamental beam, SHG fluctuations are primarily attributed to slight redistribution of energy across the spectrum (during the two-stage broadband parametric amplification in NOPA [29]) and also phase fluctuations. All experimental data presented in this contribution correspond to optimizations having passed the validity test.

Unless otherwise stated, fitness measurement for any phase profile is done by averaging the results of high enough number of laser shots to achieve an energy tolerance better (smaller) than 0.15%. The number of shots is further limited to a minimum of 200 and a maximum of 2000 shots to have practical and comparable sizes of ensembles. These parameters are only changed in experiments reported in Section 4.3.3 to assess noise robustness.

4.3. Results

Experimental verification of all investigations reported in Section 3 is a demanding (if not an impossible) task. In simulations, one optimization usually takes a few seconds (on average), and there is no uncertainty about the parameters. Experimentally, most of our optimizations take between 30 min and 2 h. Furthermore, the target phase profile may change from one lab session to another (or even during one day), and hence many results may not pass the validity test. Also, because of the stochastic nature of optimizations, the result of a valid optimization may be considerably different from those of many other optimizations performed under similar condition, as shown in Fig. 3(b). In general, while experimental results are insightful in understanding the performance of optimizations in real-life situations, they are not rigorously conclusive and should be interpreted by care.

The employed maximum values of parameters are $\tau_M = 300$ fs for delay-based optimizations and $\varphi_M = \pi$ and $\varphi_M = 20\pi$ for phase-based optimizations, based on some initial experiments. Note that the shape and the span of the estimated target phase profile in the lab are not the same as the ones considered in simulations. As such, while the trends can be similar, the specific values (optimal number and range of parameters) are not necessarily the same. The choice of $\varphi_M = \pi$ (in addition to $\varphi_M = 20\pi$) is because of its unique property in estimating phase samples over a maximal incongruent span ($\Delta\varphi = 2\pi$).

In most experiments, the (unwrapped) phase profiles found by the algorithm share a common M-shaped pattern around the center wavelength, which is primarily attributed to the non-uniformity of the phase mask. Our auto- and cross-correlation measurements of the output of the pulse shaper (with and without the phase mask) and also the input to the pulse shaper after compression by a prism compressor suggest that this M-shape pattern primarily originates from the nonuniformity of the phase mask. Note that in practice, the phase added to light by a phase mask can be decomposed into two parts; a dynamic component (as in an ideal phase mask), and also a static component originating from the nonuniformity of the phase mask (even when the control signals to all pixels are zero) [30]. What we refer to as the "input beam" in experiments is the input to the pulse shaper with the static phase of the phase mask added to it. In this way, applying a control signal to the phase mask affects the light, as an ideal phase mask would do.

Depending on the initial chirp of the input laser pulse (different from one lab session to another), the accuracy of one optimization experiment compared to another (verified by a higher fitness under similar conditions), and also the existence of a linear phase term, details of these M-shaped profiles can vary from one experiment to another (especially at wavelengths with low intensities, far from the center wavelength). The phase profiles obtained from different experiments have been added with appropriate linear (with respect to frequency) phase terms as $\varphi_{new}(\lambda) = \varphi(\lambda) + a(\lambda_0/\lambda - 1) + b$, where $\lambda_0 = 670$ nm is the center wavelength, and the fit coefficients a and b (with the common unit of radians) are chosen to match corresponding features of different phase profiles.

4.3.1. Repeatability

The learning curves and the retrieved phase profiles associated with four optimization experiments with 10 parameters are shown in Fig. 7(left) and (right), respectively. The experiments for different parameterizations are done alternatively (between the two parameterizations) to disentangle possible unwanted changes caused by laser drift. The phase-based optimization uses $\varphi_M = \pi$.

The similarity of the learning curves can be interpreted as the repeatability and the similarities of optimization trajectories, for the chosen set of parameters. The optimal solutions in Fig. 7(right) are also reasonably similar for a given parameterization. As shown in Fig. 4(bottom row), the tails of the phase profile correspond to regions with significantly smaller power spectral density, and less significant values of phase.

The smaller final fitness obtained from phase-based optimizations can simply originate from an inappropriately small value of φ_M, and does not necessarily imply the superiority of the delay-based optimization. Interestingly, the final phase profiles obtained from both parameterizations are similar in shape (but not in the range of the profiles). This observation further supports the hypothesis that a small value of φ_M has been used. In both cases, the learning curves have negative curvatures and relatively high sensitivities (fitness variations over a given number of generations) at the beginning of the optimization. The delay-based optimization features a larger sensitivity, though.

4.3.2. Range and number of parameters

The learning curves and the retrieved phase profiles associated with 6 optimization experiments are shown in Fig. 8(top) and (bottom), respectively. These data correspond to two sets of (10 and 30) parameters. For each set of parameters, we investigate the performances of optimizations with $\tau_M = 300$ fs, $\varphi_M = \pi$, and $\varphi_M = 20\pi$.

An investigation of optimizations with 10 parameters shows that all three parameterizations ($\tau_M = 300$ fs, $\varphi_M = \pi$, and $\varphi_M = 20\pi$) estimate the same shape (yet different scales) of the target phase profile. Comparing the learning curves and the phase profiles implies that the parameterization with $\varphi_M = \pi$ is using an inappropriately small range of phase, as also seen in Section 4.3.1. The similarity of final fitness values obtained by parameterizations with $\tau_M = 300$ fs and $\varphi_M = 20\pi$, despite different learning curve profiles and estimated phase profiles, suggests that final phase profiles may correspond to different local optima with similar fitness peak values in the solution landscape.

By increasing the number of parameters from 10 to 30, a small increase (~6%) is observed in the final fitness in optimizations with $\tau_M = 300$ fs and $\varphi_M = 20\pi$ parameterizations. However, the $\varphi_M = \pi$ parameterization shows a more pronounced improvement (~25%). In terms of the final estimated phase profile, the parameterizations with $\tau_M = 300$ fs and $\varphi_M = 20\pi$ show the M-shape phase pattern, but with a new small feature appeared around the center wavelength of

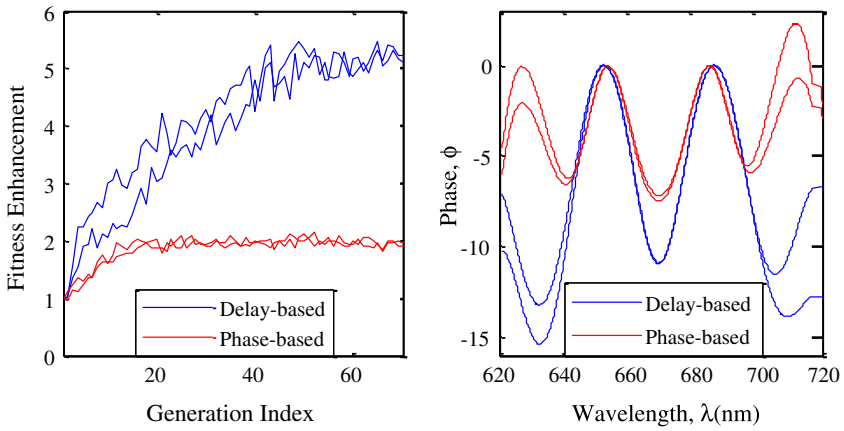

Fig. 7. (Left) The learning curves and (right) the optimal spectral phase profiles corresponding to second harmonic energy optimizations with 10 parameters: the blue (red) curves show the results obtained with delay-based (phase-based) optimizations.

670 nm. Given the slight increase of the final fitness by increasing the number of parameters (from 10 to 30), this new phase feature is considered real, and its appearance is attributed to better sampling resolution.

The impact of increasing the number of parameters from 10 to 30 is most significant in the case of the parameterization with $\varphi_M = \pi$. Comparing the blue and the red curves in Fig. 8(bottom right) shows that the second parameterization is also able to find the new feature (with relatively small amplitude) around the center wavelength of 670 nm. However, each peak of the M-shape pattern is now split into two peaks. Note that as the number of parameters increases, a profile limited between $-\pi$ and π can estimate the wrapped target profile with increasing accuracy. However, such a parameterization still has some disadvantages. It is not easy to differentiate between main features and phase wraps in the phase profile, which makes a direct and intuitive interpretation of the phase profile for spectroscopic applications difficult. Also, a constant or a linear phase term can result in a completely different-looking profile. Furthermore, increasing the number of optimization parameters slows down the algorithm.

As for the initial sensitivity, parameterizations with $\tau_M = 300$ fs and $\varphi_M = \pi$ show considerably higher sensitivities at initial stages of optimizations, compared to parameterization $\varphi_M = 20\pi$. This difference is more pronounced, as the number of parameters increases from 10 to 30. The curve of the mean fitness first drops considerably, and then increases with small slope and curvature. The slope gradually increases until reaching a turning point (around the 90th generation), after which the curvature becomes negative and the slope starts to decrease in approaching the steady state fitness. While successful in achieving high fitness values, the parameterization with $\varphi_M = 20\pi$ has a learning pattern not suitable for spectroscopic studies of samples vulnerable to precipitation or optical degradation.

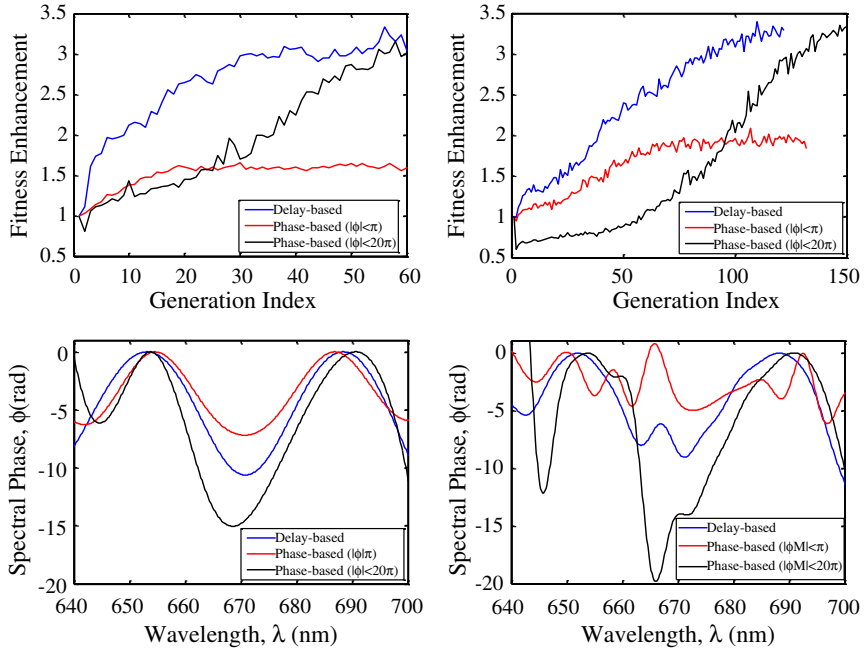

Fig. 8. (Top) The learning curves and (bottom) the optimal spectral phase profiles corresponding to second harmonic energy optimizations: the blue, the red, and the black curves correspond to parameterizations with delay, phase (limited to $\pm\pi$), and phase (limited to $\pm 20\pi$), respectively. In each row, the two plots on the left/right correspond to optimizations with 10/30 parameters. The legends show the parameterization used in each case.

4.3.3. Noise robustness

A common approach for noise reduction in pulse shaping experiments is to average out noise by increasing the number of measurements. Assuming statistical independence between the "real signal" and a contaminating additive noise, averaging the results of N measurements will decrease the noise energy by the factor N (or its amplitude by the square root of N). Here we investigate the repeatability of adaptive SHG with delay- and phase-based optimizations with different noise levels. We control the noise level indirectly by modifying the number of laser shots per measurement, as detailed in Section 4.2. The employed ranges of laser shots in these measurements are 30–60, 10–20, and 1. The experiments for different parameterizations are done alternatively with the two parameterizations (P–D–P–D–P–D …, where P and D denote phase- and delay-based experiments, respectively) to disentangle possible unwanted changes caused by laser drift.

The optimal spectral phase profiles obtained by delay-based and phase-based ($\varphi_M = 20\pi$) optimizations with 10 parameters and after 40 generations are shown in Fig. 9 with blue and red, respectively. The relative signal to noise ratio (SNR) in the measurements is symbolically encoded in the thicknesses of the curves. Fig. 9 shows that for the given choice of parameters, the delay-based optimization is more robust to noise, and the corresponding spectral phase profiles (blue curves) are more similar, compared to the profiles associated with phase-based optimization (red curves). Note that the pulse has a FWHM bandwidth of 35 nm centered at 670 nm, and the deviations of the estimated phase profiles have less significance at the tails of the spectrum.

The need for fewer (down to 1) laser shots for a successful adaptive optimization has an important practical significance, since the measurement time corresponding to each phase profile depends on the number of laser shots (on the order of hundreds). Of course, there is no need to go down to 1 laser shot for a fast optimization. The measurement time is further limited by the settling time of a phase mask (~100 ms). Also the noise suppression factor (1/N) has its highest sensitivity for small numbers of laser shots. As such, using a few laser shots can have a significant impact on the noise level, without increasing the measurement time significantly beyond the phase mask settling time.

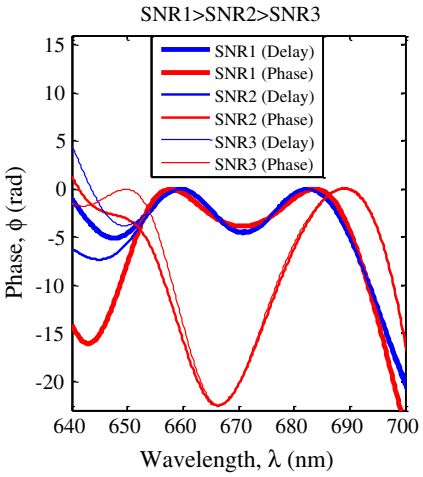

Fig. 9. The optimal spectral phase profiles obtained by second harmonic energy optimizations with different parameterizations and different numbers of laser shots: the blue/red curves correspond to parameterizations with delay/phase (limited to $\pm 20\pi$). The relative thicknesses of the curves are representatives of the number of laser shots and hence representatives of signal to noise ratio (SNR) in each case (SNR1 > SNR2 > SNR3). In all cases, 10 optimization parameters are used. The legends show the parameterization and the relative SNR used in each case.

5. Discussion

5.1. Integration, stability, and temporal shift

A time-domain integrator is an inherently unstable operator (mapping a bound constant input to an unbound ramp output) and can lead a classical control system towards instability. However, the introduced integrator in delay-based formulation is a frequency-domain integrator, and is implemented in software over a finite spectral range with no stability consequences. Also, a linear phase component (corresponding to a constant term in the delay profile) has in general no effect on a single-beam pulse shaping experiment. In experiments with multiple beams, such as pump–probe with shaped pumps [5,31], linear phase terms can be explicitly taken into account and removed to have a stationary reference point for zero delay.

5.2. Alternative (resonance) process

Here we extend our comparative study of phase- and delay-based optimizations to a different nonlinear optical process with resonance features, namely coherent anti-Stokes Raman scattering (CARS). Details of the process, the optimization, and numerical considerations have been reported elsewhere [32–34]. In brief, we study a CARS process with a broadband shaped femtosecond laser pulse centered at 800 nm with a bandwidth of 25.6 nm, and a narrowband Stokes pulse centered at 1064 nm and a bandwidth negligible compared to that of the pump (= probe) beam. For a given phase profile $\varphi(\omega)$, two CARS experiments in bulk polystyrene are performed with the phase profiles $+\varphi(\omega)$ and $-\varphi(\omega)$, and the energies of the resultant CARS signals are subtracted from each other. The fitness function is defined to be proportional to this difference. The fitness is representative of the energy of the CARS signal, by discarding the purely non-resonant component. With as few as 20 CMA-ES parameters, an optimal solution is found with a fitness value slightly greater than 2 (compared to a fitness of one by using the molecular phase profile). Increasing the number of parameters slightly changes the fitness to a maximum of 2.04, but just by minimizing interpolation errors and not by finding a new optimal phase profile [32].

The results of optimizations with phase- and delay-based parameterizations are shown in Fig. 10. Both optimizations use 25 parameters, and their learning curves are shown up to the 90th generation. They converge to close values of 2.00 (delay-based) and 2.01 (phase-based), respectively. Comparing the corresponding (mean or maximum fitness) curves shows similar trends; i.e., a sharp initial increase, followed by an increase with decreasing slope (negative curvature) towards a steady state value. However, there is an important difference between the two optimizations. The delay-

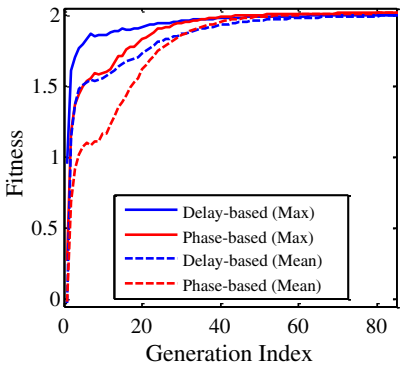

Fig. 10. The learning curves of a phase-shaped CARS process with delay-based (blue) and phase-based (red) optimization schemes. The solid/dashed curves show the maximum/mean fitness of the solution set (population) as a function of generation.

based optimization approaches significantly higher fitness values within the initial stage. In other words, a relatively high fitness value (say 1.8) is achieved by the delay-based optimization within 5 generations, while it takes the phase-based optimization more than 30 generations to approach this same value.

This initial high sensitivity of a delay-based optimization has also been observed frequently in both numerical and experimental studies of SHG optimization (see Section 4.3.1, for instance). A simple hand-waving justification is that the first population in a delay-based optimization is a (scaled) lowpass-filtered version of the first population in a comparable phase-based optimization. The required diversity for good sampling of the solution space is reasonably maintained in lowpass-filtered phase profiles. However, each individual features a more smooth profile and does not chirp the pulse dramatically. This property is expected to be more pronounced by increasing the number of parameters (with less contribution from spline smoothing in phase-based optimizations).

An interesting observation is the simultaneous improvement of both sensitivity and noise robustness by delay-based formulation, contrary to the common trade-off between the two features (as in a comparator with hysteresis or a Schmitt trigger [35]).

5.3. Parameterization and landscape subsets

A fundamental question in comparing phase- and delay-based optimizations is the similarities and differences between the optimization landscapes in these two cases. The *global* solution landscape of an optimization problem is the locus of all points in a multidimensional space, each representing a candidate solution and the associated optimization fitness. Numerical implementations of an optimization problem are commonly performed in finite-dimensional subsets of the global landscape, by imposing different constraints on the original optimization problem. While simplifying the optimization by reducing the dimensionality, such a finite-dimensional (subset) landscape can also complicate the problem by introducing new local optima or excluding the global optimum.

The landscape formed by a given differentiable basis set $\{P_n(\omega)\}$ in phase-based optimization can also be considered as the landscape corresponding to the basis set $\{dP_n(\omega)/d\omega\}$ in delay-based optimizations. In both cases, the independent variables (landscape axes) correspond to the weights of the corresponding basis functions, and the dependent variable represents the fitness. Such a landscape can also equivalently correspond to a phase- or a delay-based formulation, if the number or the ranges of landscape coordinates are limited. For example, the landscape of a phase-based optimization with the basis functions $\{\omega^2,\omega^3\}$ and weight factors limited to $[-2,3]$ can also be considered as the landscape of a delay-based optimization with basis sets $\{2\omega,3\omega^2\}$ and the same parameter range of $[-2,3]$.

In a direct parameterization (as shown in Fig. 1), the number of parameters and the type of interpolating functions (linear, cubic …) impose discrete boundary conditions and also continuous constraints between the boundaries. In this case (contrary to the aforementioned case of basis set parameterization), two "corresponding" ranges of parameters in phase- and delay-based formulations do not represent equivalent phase profiles and fitness values. Traversing optimization trajectories in such (phase- and delay-based) spaces are not expected to yield similar results, either (see Fig. 2).

Note that in a direct parameterization, the interpolating functions and hence the estimated phase function depend on the parameters in a nonlinear way, whereas in the case of basis set parameterization, the estimated phase function depends linearly (aside from a possible offset) on the landscape coordinates. Also note that practical constraints such as the finite resolution, range, sensitivity, and bit-depth of the experimental setup impose further constraints on the optimization problem and hence the solution landscape. Finally, in the presence of noise, the landscape is not stationary, and one may alternatively consider the expected value of the landscape (in a statistical sense, and different from instantaneous landscapes) [12].

6. Conclusions

Our numerical and experimental studies show an overall superiority of a delay-based formulation, compared to a conventional phase-based formulation, for adaptive laser pulse shaping. We have studied efficiency, noise robustness, convergence speed, repeatability, fitness sensitivity, and parameter insensitivity. These issues have been mostly studied in the case of second harmonic generation and briefly in the case of coherent anti-Stokes Raman scattering. The benefits of delay-based pulse shaping (for adaptive experiments) are in addition to trivial benefits of a direct and intuitive formulation, capable of efficient use of phase masks with a dynamic range $>2\pi$ (in general pulse shaping experiments). Although our results are limited to a specific choice of the range and the profile of the target phase function and the pulse spectrum, the observed trends in the results are insightful (if not the same) in more general cases.

Acknowledgments

We thank Dr. Thomas Bäck, Ing. Lars Willmes, Dr. Riccardo Fanciulli, and Dr. Peter van der Walle for their earlier help with the implementation and interpretation of the algorithm, and Ing. Niels Dijkhuizen, Ing. Jeroen Korterik, and Ing. Frans Segerink for various technical supports. This research has been supported by the *Stichting voor Fundamenteel Onderzoek der Materie* (FOM, grant numbers 03TF78-3 and 07SP01), which is supported financially by *Nederlandse Organisatie voor Wetenschappelijk Onderzoek* (NWO).

References

[1] P. Brumer, M. Shapiro, Chem. Phys. Lett. 126 (1986) 541.
[2] Y. Silberberg, Ann. Rev. Phys. Chem. 60 (2009) 277.
[3] X.G. Xu, S.O. Konorov, J.W. Hepburn, V. Milner, Nat. Phys. 4 (2008) 125.
[4] R.S. Judson, H. Rabitz, Phys. Rev. Lett. 68 (1992) 1500.
[5] J. Savolainen, R. Fanciulli, N. Dijkhuizen, A.L. Moore, J. Hauer, T. Buckup, M. Motzkus, J.L. Herek, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 7641.
[6] H. Rabitz, New J. Phys. 11 (2009) 105030.
[7] J.L. Herek, W. Wohlleben, R.J. Cogdell, D. Zeidler, M. Motzkus, Nature 417 (2002) 533.
[8] D.G. Kuroda, C.P. Singh, Z. Peng, V.D. Kleiman, Science 326 (2009) 263.
[9] J. Roslund, H. Rabitz, Phys. Rev. A 79 (2009) 053417.
[10] J. Roslund, O.M. Shir, T. Bäck, H. Rabitz, Phys. Rev. A 80 (2009) 043415.
[11] A. Jafarpour, J. Savolainen, R. de Jong, J. Middag, D.P. Sprünken, P. van der Walle, D. Yang, J.L. Herek, Opt. Express 17 (2009) 11986.
[12] P. van der Walle, H. Offerhaus, J. Herek, A. Jafarpour, Opt. Express 18 (2010) 973.
[13] A.W. Albrecht, J.D. Hybl, S.M. Gallagher Faeder, D.M. Jonas, J. Chem. Phys. 111 (1999) 10934.
[14] D. Keusters, H.S. Tan, P. O'Shea, E. Zeek, R. Trebino, W.S. Warren, J. Opt. Soc. Am. B 20 (2003) 2226.
[15] F. Langhojer, D. Cardoza, M. Baertschy, T. Weinacht, J. Chem. Phys. 122 (2005) 014102.
[16] J. Strand, T. Taxt, Appl. Opt. 38 (1999) 4333.
[17] O. Shir, Niching in Derandomized Evolution Strategies and Its Applications in Quantum Control; A Journey from Organic Diversity to Conceptual Quantum Designs, Universiteit Leiden, Leiden, 2008.
[18] R. Fanciulli, L. Willmes, J. Savolainen, P. van der Walle, T. Bäck, J.L. Herek, Lect. Notes Comput. Sci. 4926 (2008) 219.
[19] Z. Zheng, A.M. Weiner, Chem. Phys. 267 (2001) 161.
[20] N. Hansen, Tutorial: Covariance Matrix Adaptation (CMA) Evolution Strategy, Institute of Computational Science, ETH Zurich, 2006.
[21] C. Gollub, R. de Vivie-Riedle, Phys. Rev. A (Rapid Comm.) 79 (2009) 021401(R).
[22] D. Schimpf, E. Seise, J. Limpert, A. Tünnermann, Opt. Express 16 (2008) 8876.
[23] J. Vaughan, T. Feurer, K. Stone, K. Nelson, Opt. Express 14 (2006) 1314.
[24] S.D. McGrane, R.J. Scharff, M. Greenfield, D.S. Moore, New J. Phys. 11 (2009) 105047.
[25] D. Zeidler, S. Frey, K.-L. Kompa, M. Motzkus, Phys. Rev. A 64 (2001) 023420.
[26] Yuqiang Deng, Weijian Yang, Chun Zhou, Xi Wang, Jun Tao, Weipeng Kong, Zhigang Zhang, Opt. Lett. 33 (2008) 2855.
[27] Jonathan R. Birge, Franz X. Kärtner, Opt. Lett. 35 (2010) 2469.
[28] R. Chartrand, "Numerical differentiation of noisy, nonsmooth data", http://math.lanl.gov/Research/Publications/Docs/chartrand-2010-numerical.pdf, http://math.lanl.gov/Research/Publications/chartrand-2010-numerical.shtml, Submitted, 2010.
[29] S. Lochbrunner, T. Wilhelm, J. Piel, S. Spörlein, E. Riedle, OSA Trends Opt. Photon. Ser. 26 (1999) 366.

[30] X. Xun, R.W. Cohn, Appl. Opt. 43 (2004) 6400.
[31] A.L. Dobryakov, J.L. Perez Lustres, S.A. Kovalenko, N.P. Ernsting, Chem. Phys. 347 (2008) 127.
[32] A.C.W. van Rhijn, H.L. Offerhaus, P. van der Walle, J.L. Herek, A. Jafarpour, Opt. Express 18 (2010) 2695.
[33] S. Postma, A.C.W. van Rhijn, J.P. Korterik, P. Gross, J.L. Herek, H.L. Offerhaus, Opt. Express 16 (2008) 7985.
[34] A.C.W. van Rhijn, S. Postma, J.P. Korterik, J.L. Herek, H.L. Offerhaus, J. Opt. Soc. Am. B 26 (2009) 559.
[35] Otto H. Schmitt, J. Sci. Instrum. 15 (1938) 24.