Robust orthogonal parameterization of evolution strategy for adaptive laser pulse shaping

Aliakbar Jafarpour*, Janne Savolainen, Rianne de Jong, Jacob Middag, Daan P. Sprünken, Peter van der Walle, Di Yang, and Jennifer L. Herek

Optical Sciences group, MESA + Institute for Nanotechnology, University of Twente, the Netherlands
*jafarpoura@gmail.com

Abstract: Many spectroscopic applications of femtosecond laser pulses require properly-shaped spectral phase profiles. The optimal phase profile can be programmed on the pulse by adaptive pulse shaping. A promising optimization algorithm for such adaptive experiments is evolution strategy (ES). Here, we report a four fold increase in the rate of convergence and ten percent increase in the final yield of the optimization, compared to the direct parameterization approach, by using a new version of ES in combination with Legendre polynomials and frequency-resolved detection. Such a fast learning rate is of paramount importance in spectroscopy for reducing the artifacts of laser drift, optical degradation, and precipitation.

References and Links

1. P. Brumer, and M. Shapiro, "Control of unimolecular reactions using coherent light," Chem. Phys. Lett. **126**(6), 541–546 (1986).
2. S. Shi, A. Woody, and H. Rabitz, "Optimal control of selective vibrational excitation in harmonic linear chain molecules," J. Chem. Phys. **88**(11), 6870–6883 (1988).
3. J. Werschnik, and E. K. U. Gross, "Quantum optimal control theory," J. Phys. At. Mol. Opt. Phys. **40**(18), R175–R211 (2007).
4. R. S. Judson, and H. Rabitz, "Teaching lasers to control molecules," Phys. Rev. Lett. **68**(10), 1500–1503 (1992).
5. B. J. Pearson, J. L. White, T. C. Weinacht, and P. H. Bucksbaum, "Coherent control using adaptive learning algorithms," Phys. Rev. A **63**(6), 063412–12 (2001).
6. J. L. Herek, W. Wohlleben, R. J. Cogdell, D. Zeidler, and M. Motzkus, "Quantum control of energy flow in light harvesting," Nature **417**(6888), 533–535 (2002).
7. J. Savolainen, R. Fanciulli, N. Dijkhuizen, A. L. Moore, J. Hauer, T. Buckup, M. Motzkus, and J. L. Herek, "Controlling the efficiency of an artificial light-harvesting complex," Proc. Natl. Acad. Sci. U.S.A. **105**(22), 7641–7646 (2008).
8. C. Riziotis, and A. V. Vasilakos, "Computational intelligence in photonics technology and optical networks: A survey and future perspectives," Inf. Sci. **177**(23), 5292–5315 (2007).
9. A. Assion, T. Baumert, M. Bergt, T. Brixner, B. Kiefer, V. Seyfried V, M. Strehle, and G. Gerber, "Control of chemical reactions by feedback-optimized phase-shaped femtosecond laser pulses," Science **282**(5390), 919–922 (1998).
10. H.-G. Beyer, "Evolutionary algorithms in noisy environments: theoretical issues and guidelines for practice," Comput. Methods Appl. Mech. Eng. **186**(2-4), 239–267 (2000).
11. N. Hansen, "The CMA evolution strategy: a tutorial," http://www.lri.fr/~hansen/cmatutorial.pdf.
12. D. Zeidler, S. Frey, K.-L. Kompa, and M. Motzkus, "Evolutionary algorithms and their application to optimal control studies," Phys. Rev. A **64**(2), 023420–13 (2001).
13. O. Shir, *Niching in derandomized evolution strategies and its applications in quantum control; a journey from organic diversity to conceptual quantum designs* (Universiteit Leiden 2008).
14. Ch. Siedschlaga, O. M. Shirb, Th. Bäckb, and M. J. J. Vrakkinga, "Evolutionary algorithms in the optimization of dynamic molecular alignment," Opt. Commun. **264**(2), 511–518 (2006).
15. J. W. Wilson, P. Schlup, M. Lunacek, D. Whitley, and R. A. Bartels, "Calibration of liquid crystal ultrafast pulse shaper with common-path spectral interferometry and application to coherent control with a covariance matrix adaptation evolutionary strategy," Rev. Sci. Instrum. **79**(3), 033103 (2008).
16. D. Cardoza, F. Langhojer, C. Trallero-Herrero, O. L. A. Monti, and T. Weinacht, "Changing pulse-shape basis for molecular learning control," Phys. Rev. A **70**(5), 053406 (2004).
17. R. Fanciulli, L. Willmes, J. Savolainen, P. van der Walle, T. Bäck, and J. L. Herek, "Evolution strategies for laser pulse compression," Lect. Notes Comput. Sci. **4926**, 219–230 (2008).

18. A. M. Weiner, "Femtosecond pulse shaping using spatial light modulators," Rev. Sci. Instrum. **71**(5), 1929–1960 (2000).
19. E. Barker and J. Kelsey, *Recommendation for Random Number Generation Using Deterministic Random Bit Generators* (NIST Special Publication 800–90 2007).
20. A. Ostermeier, A. Gawelczyk, and N. Hansen, "Step-size adaptation based on non-local use of selection information," in *Parallel Problem Solving from Nature–PPSN III*, G. Goos, J. Hartmanis, and J. van Leeuwen, eds. (Springer 1994), pp. 189–198.
21. N. Hansen, and A. Ostermeier, "Completely derandomized self-adaptation in evolution strategies," Evol. Comput. **9**(2), 159–195 (2001).
22. D. Meshulach, D. Yelin, and Y. Silberberg, "Adaptive ultrashort pulse compression and shaping," Opt. Commun. **138**(4-6), 345–348 (1997).
23. Z. Zheng, and A. M. Weiner, "Coherent control of second harmonic generation using spectrally phase coded femtosecond waveforms," Chem. Phys. **267**(1-3), 161–171 (2001).
24. J. Roslund, M. Roth, and H. Rabitz, "Laboratory observation of quantum control level sets," Phys. Rev. A **74**(4), 043414–11 (2006).
25. A. B. Carlson, *Communication Systems* (McGrawHill, 1986).
26. B. Mulgrew, "Orthonormal functions for nonlinear signal processing and adaptive filtering," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. **6**, (ICASSP 1994), pp.509–512.
27. B. Ninness, S. Gibson, and S. Weller, "Practical aspects of using orthonormal system parameterisations in estimation problems," in *Proc. 12th IFAC Symposium on System Identification*, vol. **2**, B. Ninness and H. Hjalmarsson ed., (SYSID 2000), pp. 463–468.
28. B. J. Pearson, J. L. White, T. C. Weinacht, and P. H. Bucksbaum, "Coherent control using adaptive learning algorithms," Phys. Rev. A **63**(6), 063412 (2001).

1. Introduction

Femtosecond laser pulses are widely used in spectroscopy for monitoring or controlling electronic, excitonic, plasmonic, and vibrational transitions. Finding a laser pulse to optimize an ultrafast physical or chemical process [1] is a classical problem in optimal control theory [2,3]. Once the dynamics (Hamiltonian) of a system are known, the control problem can be formulated. However, analytical or even numerical solutions are only available for simple systems. In most cases, an adaptive scheme is used to find the optimal laser pulse [4–7]. A pulse shaper tailors the properties of the input pulse, and an optimization algorithm determines the next pulse shape in accordance with a feedback signal from the optical process, and this process is iterated (Fig. 1).

Evolutionary algorithms such as genetic algorithm and evolution strategy (ES) are commonly used in optics [8], and especially in adaptive laser pulse shaping [9,10]. These stochastic algorithms search for optimal solutions in an iterative optimization process. The optimization is based on the computational intelligence of the algorithm and finding patterns in numerical data, rather than an analytic formulation of the problem or the control strategy. Recently, an advanced ES with an improved adaptation mechanism, known as covariance matrix adaptation (CMA-ES) [11], has shown promise for adaptive laser pulse shaping.

Previous studies of the applicability of ES for adaptive pulse shaping are limited, especially in speed of convergence and robustness to noise. Zeidler et al. reported a detailed numerical and experimental study of ES and investigated the practical use of such algorithms for adaptive pulse shaping [12]. However, an older version of ES was used, only one type of noise was considered, orthogonal basis functions were not employed, and a photomultiplier tube was used along with a boxcar system to reduce the overall effect of noise by filtering the signal in the time domain. Shir et al. reported a comprehensive numerical study of the CMA-ES for a few different adaptive pulse shaping test cases, along with some experiments using sensitive box-car detection [13]. They considered orthogonal basis functions only in simulations with a quadratic input phase profile in the absence of noise. Siedschlag et al. reported a comparative study of different evolutionary algorithms in the study of dynamic molecular alignment by shaped femtosecond pulses [14]. Wilson et al. showed the efficiency of CMA-ES compared to the original ES in a second harmonic generation experiment for a population with a relatively large number of parameters [15]. Cardoza et al. reported a reduction of control parameters to one degree of freedom by appropriate choice of the basis functions in a molecular fragmentation learning control experiment [16]. The first

comprehensive experimental and numerical study of the role of noise in adaptive pulse shaping with CMA-ES was reported by Fanciulli et al [17].

An important issue in all previous studies of CMA-ES was the need for a relatively large number of iterations to converge to a solution. Here, we report the experimental result of a four-fold increase in the rate of convergence of CMA-ES (the inverse of the number of generations required for convergence) and an improved (by 10%) final result compared to what was achieved before. Furthermore, the effects of four types of noise are studied to model both detection noise and laser instability. Frequency-resolved detection is used to create more robustness to noise with no need for sensitive detection schemes. Finally, we investigate and exemplify the effect of parameterization on the apparent shape of the landscape.

The ES algorithm and different aspects of adaptive pulse shaping are described in Section 2. Simulation and experimental results are shown in Sections 3 and 4, respectively. Section 5 presents a detailed discussion of the results, and final conclusions will be made in Section 6.

2. Concepts and methodologies

2.1 Adaptive laser pulse shaping

Figure 1 shows a block diagram of an adaptive learning loop for laser pulse shaping. A controller (pulse shaper) changes the properties of the input laser pulse before applying it to the optical process. An optimization algorithm uses feedback from the experiment to generate a new set of parameters, which change the properties of the pulse shaper via a chosen mapping (parameterization). The arrow crossing the pulse shaper signifies the adaptive nature of the control, in which feedback modifies the controller itself and not the input to the controller. The optical dynamics in the feedforward direction occurs at the femtosecond time scale, whereas the (opto) electronic dynamics in the feedback path, mainly limited by the integration time of the detector and the settling time of the pulse shaper, occur at a much longer time scale (~0.1s). Laser instability and detection errors are modeled as input and output noise, respectively.

Here, we consider pulse shaping schemes that modify only the spectral phase profile of a linearly polarized input laser pulse; and the amplitude and polarization are left unchanged. In practice, a pixelated "phase-only" pulse shaper can also affect the amplitude because of diffraction [18]. We neglect this minor effect that does not affect the comparison of different formulations of the optimization problem.
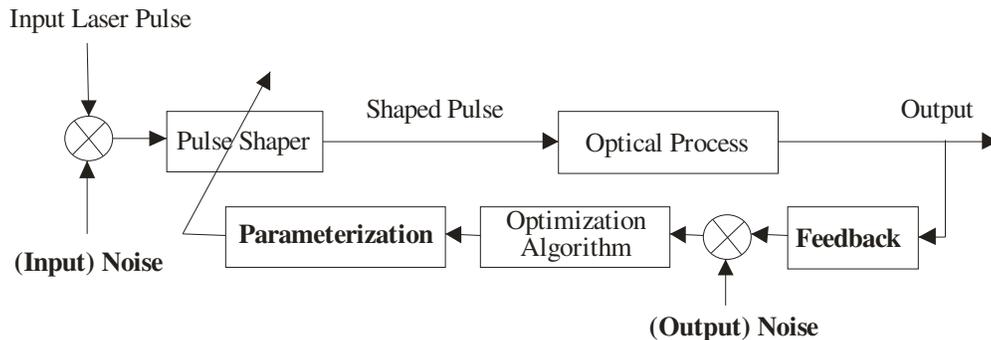

Fig. 1. A model block diagram for an adaptive laser pulse shaper.

2.2 Evolution strategy

Evolutionary algorithms are stochastic blind optimization techniques that seek a set or population of solutions at the same time. Using a population instead of a single solution, results in exploring different directions in the solution space (fitness landscape) and a better chance of coming out of local optima and saddle points in the landscape. Different runs of the same stochastic optimization problem (with different seeds for pseudorandom number

generators [19]) do not necessarily generate similar results, even in the absence of noise. Hence, there is always a distribution of solutions, and the average of different runs is considered "the solution".

An important property of evolutionary algorithms is the spread (or variance) of the population. A small spread near the solution results in a fast learning, and a large spread at the beginning results in exploring different directions in the landscape. However, an inappropriate spread results in slow convergence or incomplete exploration of the landscape. An evolution strategy tackles this issue by dynamically adapting the step size that mutates the population [20]. CMA-ES is a recent powerful version of ES using the correlations in the population by adapting the covariance of the mutations to the problem [21].

Contributions of this paper concern the bold-labeled components in Fig. 1. The focus of this study is the effect of parameterization on the optimization process and its robustness to noise. The implementation used here has been done with the best CMA-ES parameter settings found in Ref. 14. The reader is referred to Ref. 11 for details of CMA-ES.

2.3 Second harmonic generation as a test case

We consider collinear type I phase-matched second harmonic generation (SHG) of a laser pulse in the non-depleted pump regime. The pulse is incident normally on a "thin" uniaxial crystal with instantaneous χ^2 nonlinearity and ideal anti-reflection coatings on both sides. This SHG signal is the result of both second harmonic and sum-frequency generation of individual frequencies within the broadband pulse and is affected by group properties of the pulse. The full-width half maximum (FWHM) bandwidth of the crystal phase-matching profile is considerably larger than the FWHM bandwidth of the laser pulse spectrum. Avoiding tight focusing and using a thin crystal guarantee the validity of a simple plane wave model for the SHG process and negligible walk-off and transverse modulations ("rings") of the nonlinear beam. Under these conditions and with the slowly-varying envelope approximation (SVEA), not only the nonlinear polarization, but also the nonlinear component of the electric field is proportional to the square of the input electric field in the time domain.

Experiments [22] and analyses [23] show that maximizing the total energy of the SHG signal from a thin crystal is equivalent to achieving a flat spectral phase profile. Hence, SHG optimization is commonly used in pulse shaping spectroscopy as a preliminary step to achieve a flat-phase pulse at the sample position before shaping the pulse for the main experiment. It is also used as an independent test case to evaluate an adaptive shaping system.

2.4 The fitness landscape in SHG optimizations

The performance of an adaptive pulse shaping experiment can be understood better by having some insight into the shape of the landscape. Geometric and topological parameters such as the gradient, curvature, torsion, orientability, connectivity, and extension of the landscape affect optimization features such as the rate of learning, the convexity of the optimization, and robustness to noise.

Assuming that an analytic function describes the unwrapped spectral phase, i.e., the integral of the group delay, the phase function can be expanded in a Taylor series around the center frequency. By further assuming SVEA and discarding delay, we will neglect the zero order and first order terms of the series. In this study, any phase function will be modeled by a few coefficients of this Taylor expansion, starting from the second order term.

Figure 2 (left) shows SHG spectra, to which higher orders of spectral phase are applied for a laser pulse with a Gaussian spectrum centered at 800nm with a FWHM bandwidth of 35nm. While all these different orders of phase have a similar effect in reducing the total energy of the SHG pulse by 50%, they result in SHG profiles with different slopes, curvatures, and extensions. The SHG spectra are all symmetric in the frequency domain and have the same peak value. So, any decrease in the peak value or any asymmetry in the spectral profile of the SHG signal (for pulses such as the input pulse considered in this case) implies simultaneous presence of multiple phase orders.

The amount by which the SHG energy drops as a result of introducing a given spectral phase order has been depicted in Fig. 2 (right). In order to have a meaningful comparison between different phase orders, we consider a range of each phase term, over which the SHG energy is reduced to half of the SHG energy of the transform limited (TL) pulse. As the phase order increases, the magnitudes of the slope and curvature at the center of the normalized energy traces increase, while they decrease at the tails. The maximum values of these three phase orders are $\varphi_{2,max} = 453fs^2$, $\varphi_{3,max} = 26,000fs^3$, and $\varphi_{4,max} = 785,000fs^4$.
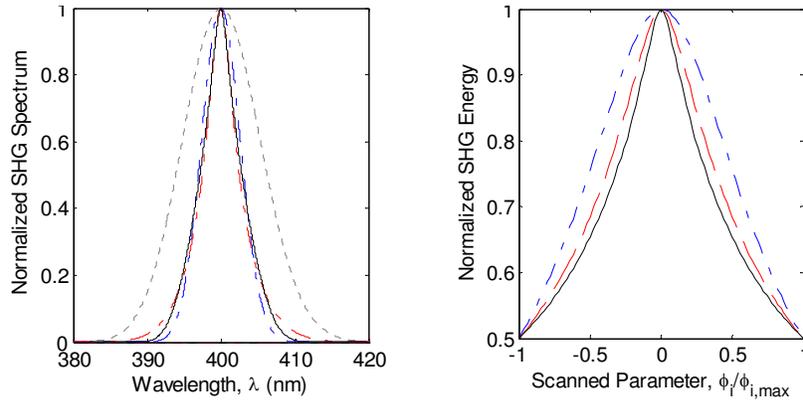

Fig. 2. (Left) Second harmonic spectrum of a transform-limited Gaussian pulse centered at 800nm with a bandwidth of 35nm (dot), along with second harmonic spectra of the same pulse with second order (dash-dot), third order (dash), and fourth order (solid) phase terms. (Right) Variations of the total second harmonic energy as the second order (dash-dot), third order (dash), and fourth order (solid) phase terms are scanned one at a time.

The effect of simultaneous presence of two phase orders in the input laser pulse is shown in Fig. 3 (left). The combination of cubic (φ_3) and quartic (φ_4) phase terms results in a more concave pattern compared to the combination of φ_3 and φ_2. The combination of φ_2 and φ_4 gives rise to considerably different effects compared to the combination of these even-order terms with φ_3. There are no reflection symmetries, contours of constant SHG energy form a twisted pattern, and values of φ_2 and φ_4 with opposite signs cancel each other considerably. The twisted shape of the contours will result in a spiral trajectory in a steepest-descent optimization.

(a)

Contours of Equal SHG Energy for different phase terms
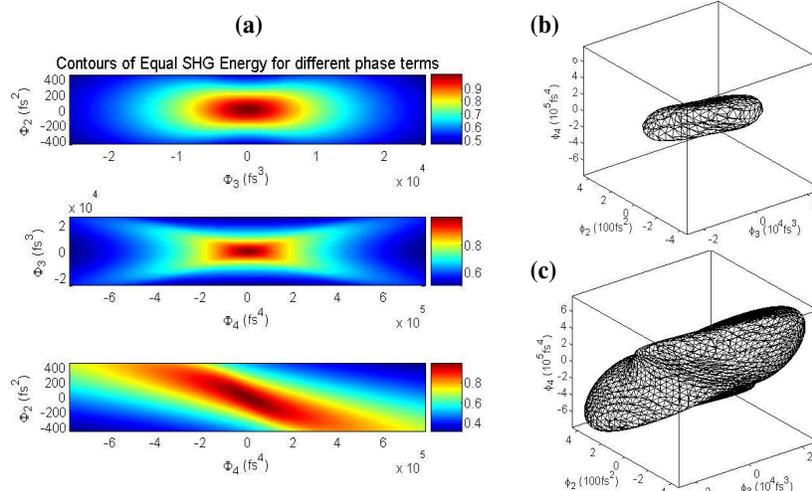
(b)

(c)

Fig. 3. Variations of the total second harmonic energy of a laser pulse (a) when two phase orders exist simultaneously in the input phase profile. (b) The level set of the total second harmonic energy generated by the second, the third, and the fourth order phase terms for a normalized energy of 0.8 and (c) 0.65.

The symmetries of 2D contours of SHG energy shown in Fig. 3(left) imply that the SHG energy depends only on the magnitude and the relative signs of different phase terms. The common point symmetry of all three contours originates from the time-reversal ambiguity of the SHG process. Similar conclusions can be made by considering 3D level sets for a few different values of SHG energy shown in Figs. 3(b,c). An SHG level set is the locus of all points in the $(\varphi_2, \varphi_3, \varphi_4)$ space with equal SHG energy. As the energy decreases, the torsion of the level sets becomes more visible, and the level sets are elongated along a line with a negative slope in the φ_2-φ_4 plane, since opposite signs of φ_2 and φ_4 partially cancel each other. A perturbative analysis [24] shows that for nearly transform-limited Gaussian pulses, the 3D level sets are expressed by quadratic equations with positive terms; i.e., they are ellipsoids around a global maximum at the origin, and the optimization is convex.

3. CMA-ES simulations

3.1 Input phase profile

In order to do a realistic simulation of experiments, we consider the input phase profile in our simulations to be the phase profile found in previous SHG optimizations, as shown in Fig. 4. Even after dismantling and setting up the measurement setup or full realignment of the laser source, the SHG optimization has always resulted in such a phase profile. The cause of this profile is mainly the phase mismatch in a parametric amplification process and the variation in thickness of the liquid crystal mask in the pulse shaper. The phase profile for the simulations has been obtained from the experimental solution by unwrapping and applying a fifth order polynomial fit to the data. The large variations at the edges of the profile are less accurate, since they are associated with the tails of the spectrum. Experiments show that extrapolating or blanking the phase in or beyond these regions does not have any noticeable impact on the second harmonic spectrum, since the signal energy is comparable or lower than that of noise. The input spectrum is a Gaussian with a bandwidth of 30nm centered at 670nm.
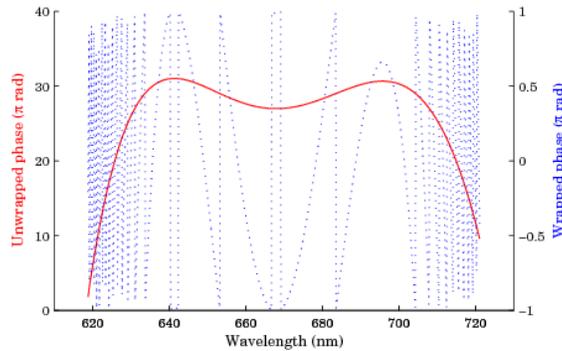
Fig. 4. A typical phase profile found using our previous optimization experiments and used here as the phase of the input laser pulse in our simulations.

3.2 Numerical settings and considerations

In modeling spectral phase, the frequency variable is centered and scaled (zero mean and unity maximum) for numerical stability and robustness. Our previous experiments [17] used a pulse shaper with 640 pixels and a dispersion coefficient of 0.16nm/pixel. Here in our simulations, we consider a larger spectral span (better temporal resolution) and higher spectral resolution (larger temporal range) to have a better modeling accuracy compared to the shaping accuracy. The number of data points has been rounded to 4096 to use the fast Fourier transform (FFT) efficiently in the SHG simulations. The CMA-ES algorithm is used with a population size of 40 with 20 parents and an initial step size of 10% [17]. The number of generations (iterations) is limited to 150, corresponding to a maximum of two hours for experimental optimizations. With any set of parameters, the simulation program is run twenty times to find the average solution.

3.3 Tested basis functions

We compare the performance of polynomial basis functions, Fourier series, and a few different orthogonal basis functions as the parameterization (see Fig. 1). Our preliminary simulations have shown a better performance of Legendre and (first kind) Chebyshev functions compared to other orthogonal functions. These two basis functions are both from the category of orthogonal basis functions defined on a finite interval (Jacobi-like functions, as opposed to Laguerre-like and Hermite-like functions). Modeling a laser pulse with a finite spectral content by using a Jacobi-like function results in a linear mapping of the pulse spectrum to the input variable of the Jacobi-like function. However, modeling such a laser pulse with the two other types of orthogonal functions results in a nonlinear mapping of the pulse spectrum to the input variable of these functions. Such nonlinear mappings change the shape and the curvature of the fitness landscape and can negatively affect the performance of the optimization algorithm. Given the promise of our preliminary simulations with Jacobi-like functions, we restrict the use of orthogonal functions to Legendre and Chebyshev functions in the rest of this study. A few different orders of the four basis functions used in the following simulations are shown in Fig. 5.
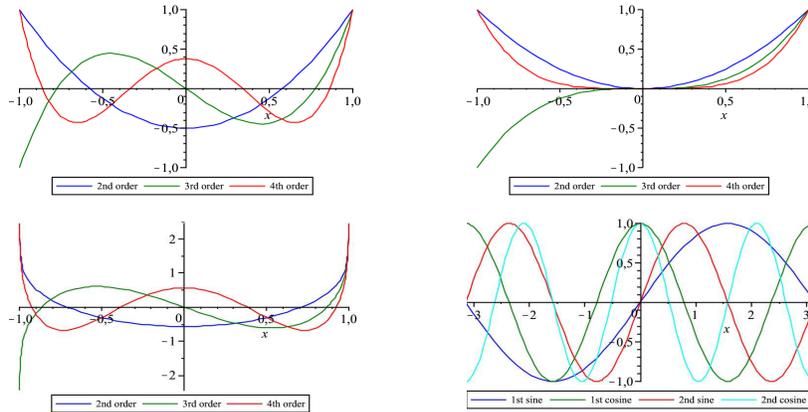
Fig. 5. The first few orders of the four basis functions used in our simulations: (top left) Legendre, (top right), Polynomial, (bottom left) Chebyshev, and (bottom right) Fourier basis functions

3.4 Phase scale and number of parameters

In each iteration (generation), the algorithm generates 40 vectors (population size). The elements (parameters) of each vector are the coefficients of a basis function set. The number and the range of variation of these parameters are two important attributes of an optimization. A combination of different orders of a basis function should be scaled, so that the modeled range is large enough to match the range of variation of a given phase function without taking the algorithm several generations to narrow down the search to the right subspace. A single scale factor can be used for all orders of a basis function, if different orders are first properly pre-scaled with respect to each other. As will be shown later, polynomial and Legendre functions defined on the interval [-1,1] have shown excellent performance just with one scale factor for all orders with no need to a pre-scaling array.

In a polynomial (or similar) fit, the order of the fitting function should be high enough to model small variations in a data set, and low enough to avoid ill-conditioned matrices or other numerical errors, which are more problematic in the presence of noise. A "small" error in the estimation of a higher order coefficient can lead to large errors, when multiplied by large numbers (higher exponents). A common solution is to center and scale data sets to zero mean and unity variance.

The effect of the aforementioned scaling and the number of parameters on the learning process in the absence of noise was evaluated in repeated simulations. Figure 6 shows the variations of the mean fitness of the SHG optimization simulation after 150 generations for Legendre and polynomial basis functions. These results correspond to different values of the scale factor and the basis function order. Both basis functions achieve a high fitness close to unity, for relatively large ranges of the scale factor (30 to 80 for Legendre and 150 to 240 for polynomial parameterization) and considerable robustness to the choice of the order (5 to 8). The input phase function in all these cases is a fifth-order polynomial.

A good optimization will be obtained when the right order is chosen. A smaller order will result in a smaller fitness, but there is some robustness to the choice of an estimation order greater than the actual order of the function to be estimated. This conclusion has also been confirmed by our experiments. Using the information in Fig. 6 and preliminary experimental results, we choose a dimension of 6 in all experiments reported here. Note that these parameterizations utilize a scaled and centered variable to model the shaping coordinate. When the direct frequency is used as a parameter, the sensitivity to the choice of the order will increase.
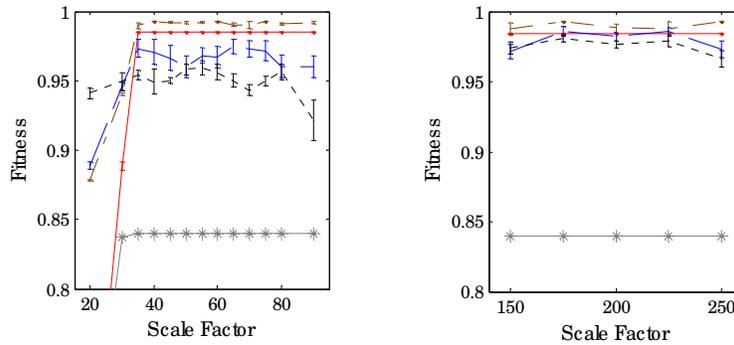
Fig. 6. Variations of the mean fitness found after 150 generations using (left) Legendre and (right) polynomial basis functions. In each case, five sets of simulations have been done with the maximum order of each basis function being four (star), five (solid), six (dash-dot), seven (dash), and eight (dot).

3.5 Noise

As an evolutionary algorithm approaches an optimal solution, the population concentrates around the optimal solution point in the fitness landscape. Hence, the plots of the minimum, the mean, and the maximum fitness as functions of generation all converge to the same point. The presence of noise has two salient features; the separation of these three curves and their convergence to smaller fitness values.

Using simulations, we investigate the influence of the noise level on the final fitness of a parameterization. The final fitness values achieved by different parameterizations and their robustness to noise after 150 generations are evaluated. Noise sources are limited to four: an additive and a multiplicative source at the input (affecting the input laser pulses) to model laser instability, and also an additive and a multiplicative source at the output (affecting the detected signal) to model inherent electronic and optoelectronic fluctuations superimposed on the detected signal per pixel of the spectrometer. The latter group of noise sources also models the discretization noise of the embedded analog to digital converter of the spectrometer, stray light fluctuation, and electromagnetic interferences.

We assume real noise sources that vary with time, but their statistical properties such as power or histogram remain constant (stationary noise). A nearly uniform power spectral density (white noise) with a Gaussian distribution is considered for all sources. In case of output noise sources, we consider the same noise statistics for a snapshot of the spectrum and for fluctuations of a frequency component over time (ergodic noise) [25].

Our goal is to compare the robustness of different parameterizations to the simultaneous presence of all noise sources. We first ran a series of simulations to find the range of signal to noise ratio (SNR) for each noise source that affects the learning process without stopping it. Then, an arbitrary set of such SNR values were chosen to be used in all simulations to compare different parameterizations. These arbitrary values are SNR = 14, 26, 7, and 12 dB for the additive input, multiplicative input, additive output, and the multiplicative output noise sources, respectively. Different relations between the statistical and physical energies of input and output signals and also different spans of signals have been considered in the calculation of SNR values.

3.6 Comparative study of different parameterizations

The common technique in estimating the value of phase at each pixel of a pulses shaper is to assign the outputs of the algorithm directly to that pixel. In practice, a fraction of the number of pixels is used to model a decimated phase profile. Interpolation is then used to calculate the values of the phase function at other pixels. We refer to this parameterization (with or without interpolation) as direct parameterization. Alternatively, a basis function set can be used to

model the spectral phase function, and the outputs of the algorithm only determine the coefficients of different orders of that basis function.

A good value for the number of pixels for interpolation based on our previous experiments was found to be 192 (compared to 640 physical pixels). However, parameterization using a basis function requires significantly fewer (< ~10) number of parameters. Another advantage of using basis functions for parameterization is the possibility of estimating the unwrapped phase function. Stable direct parameterization [17], however, converges to a wrapped phase profile, and it makes the step size estimation by the algorithm more vulnerable to the congruent nature (modulo 2π) of the phase function.
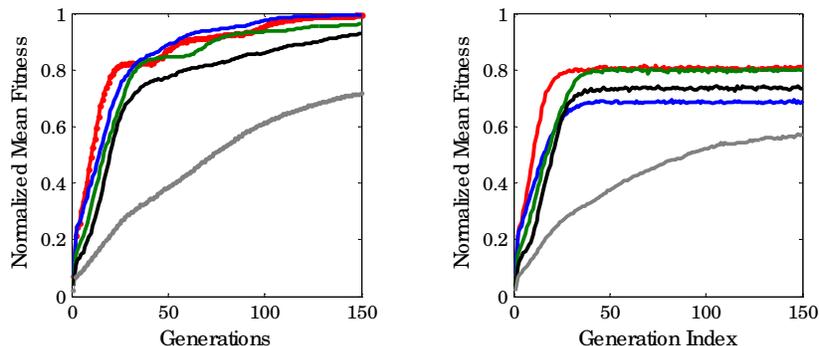

Fig. 7. The effect of different parameterizations on the learning dynamics of the algorithm in the absence (left) and presence (right) of noise. The employed parameterizations are direct (grey), Fourier series (black), Chebyshev (green), polynomial (blue), and Legendre (red) functions.

The learning process of the algorithm with different parameterizations is shown in Fig. 7. All basis functions are considerably faster than direct parameterization, both in the presence and absence of noise. Orthogonal basis functions have the same ranking with or without noise. While the polynomial basis function outperforms all orthogonal basis functions in the absence of noise, it shows the lowest fitness value among basis functions in the presence of noise. The final fitness found in all cases has decreased in the presence of noise. The fastest learning and the highest fitness correspond to Legendre functions in both cases.

4. Experiments

To test and verify the applicability of aforementioned parameterizations in practical laser pulse shaping applications, we use the output of a femtosecond laser amplifier (Legend Elite, Coherent Inc.) generating pulses with a FWHM bandwidth of 35nm centered at 800nm. The pulse shaper uses a 2D spatial light modulator (HEO1080P, Holoeye Corp.) with 1920x1080 pixels used in a reflective 4-f zero-dispersion geometry. The output of the pulse shaper has a beam diameter of 3mm, and is focused onto a 10-micron thick BBO crystal using a lens with a focal length of 200mm. The pulse energy is set with an adjustable filter before the shaper to operate in the non-depleted pump regime, while having a reasonable SNR. SHG pulses are then collimated and free-space coupled to a visible-range fiber spectrometer (AvaSpec-3648-USB2, Avantes BV). The total SHG energy is calculated by integrating the SHG spectrum.

The learning curves of adaptive SHG optimization with Legendre and interpolation parameterizations are shown in Fig. 8. Using the Legendre basis function results in a significant (more than 4-fold) improvement in the rate of learning. It also results in a fitness value more than 10% greater than that achieved by direct optimization. Fitness values have been normalized to the fitness value of the best shaped pulse. Figure 8 also shows the retrieved spectral phase functions using each parameterization. Note the piecewise linear (multiple slopes) of the learning curve for interpolation in Fig. 8 (left). This behavior is observed frequently in optimization experiments, and implies traversing a new trajectory in

the landscape. Figure 8 (right) shows the retrieved spectral phase functions by the two parameterizations.

In practice, coefficients of different Legendre functions only used 10% of the resulting range. This limited range means that it has taken the algorithm a few generations to narrow down a large search space to a smaller subset and the convergence could be even faster using smaller scale factors. This behavior and the relative insensitivity to the range are important factors and agree with the simulation results in Fig. 6.
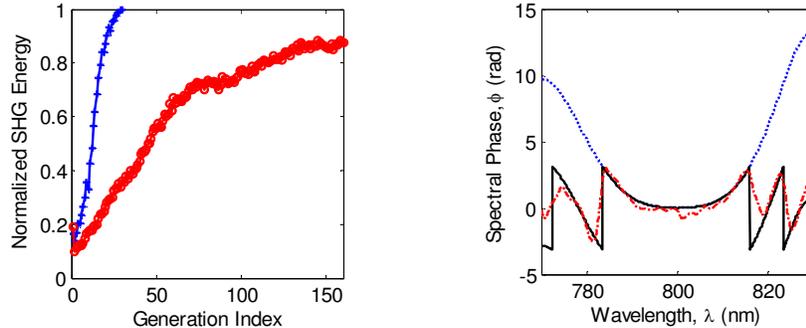

Fig. 8. Experimental results of using different parameterizations for adaptive second harmonic generation: (Left) Fitness vs. generation curves for Legendre (cross) and direct (circle) parameterizations, and (Right) the retrieved spectral phase profiles using Legendre (dot) and direct (dash-dot) parameterizations. The phase profile retrieved using Legendre parameterization is also shown in the wrapped form (solid) to be comparable with the profile obtained by direct parameterization.

The results of SHG optimization using Legendre and polynomial basis functions are shown in Fig. 9. An input phase profile with the opposite curvature and a larger span, compared to previous experiments (Fig. 8) was chosen for these experiments. Scale factors of 50 and 250 were chosen for the Legendre and polynomial basis functions, respectively. The results show that optimizations with Legendre basis functions are faster and eventually converge to a slightly higher value compared to those with polynomial basis functions, in agreement with the results in Fig. 7. Since both basis functions estimate the unwrapped phase and converge to values close to each other, they estimate similar spectral phase functions over the spectral range, where the input beam has a relatively high energy.

In both sets of measurements (Fig. 8 and 9), each optimization has been done three times, and the average of three optimization experiments has been shown here for each parameterization. The relative performance of two parameterization schemes in each pair of experiments is similar to the relative performances of averaged results. Because of the small difference between the performances of polynomials and Legendre functions (Fig. 9), our experiments with the two parameterizations were performed in an alternating fashion to suppress the effects of possible drifts and disturbances.
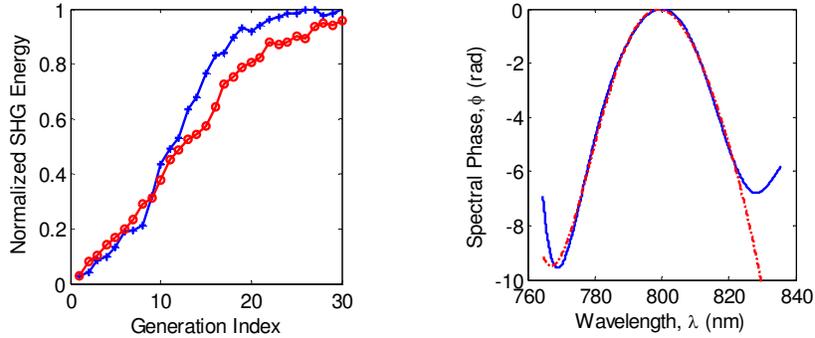
Fig. 9. Experimental results of using different parameterizations for adaptive second harmonic generation: (Left) Fitness vs. generation curves for Legendre (cross) and polynomial (circle) parameterizations, and (Right) the retrieved spectral phase profiles using Legendre (solid) and polynomial (dash-dot) parameterizations.

5. Discussion

We have successfully reduced the number of generations required for the CMA-ES algorithm to converge to an optimal solution of the SHG optimization problem. Experimental results shown in Fig. 8 and 9 confirm the behavior predicted by simulation results shown in Fig. 7(b). In this Section, we address a few aspects of the improvements in adaptive pulse shaping, reported here.

5.1 Orthogonal parameterization vs. orthogonal parameters

An important question is whether the use of an orthogonal basis function for parameterization results in an orthogonal set of phase parameters; i.e., if the optimal values of different phase terms can be found independently and by scanning one parameter at a time. The short answer is negative. The main mechanism of orthogonalizing phase parameters in a CMA-ES optimization is the inherent and continuous rotation of the vector of parameters by the covariance matrix [11]. It can also be understood by observing the results in Fig. 7(a), in which the non-orthogonal polynomial basis function is as good as orthogonal basis functions. Note that although polynomials are not orthogonal, they are independent basis of a vector space.

In order to understand the effect of parameterization with different functions on the SHG landscape, we go back to the twisted 2D contours shown in Fig. 3, with $(\varphi_2, \varphi_4) = (1,0)$ and $(\varphi_2, \varphi_4) = (0,1)$ being the principal parameter axes. Our goal here is to get insight into the changes of the fitness landscape caused by a given parameterization via a numerical study. An analytical treatment of the problem can utilize an approach such as the perturbative one detailed in Ref. 21.

We investigate the effect of transforming the (φ_2, φ_4) parameter coordinate by changing the fourth order polynomial from ω^4 to $\omega^4 - \eta\omega^2$ to disentangle the coupling between φ_2 and φ_4. Here, ω is the centered and scaled radian frequency, and η is a constant. The evolution of the contour plots, when the fourth-order polynomial is changed to a generalized fourth-order function is shown in Fig. 10. All other simulation parameters are similar to those for simulations the results of which are shown in Figs. 2 and 3. A second change of replacing the second-order polynomial with $\omega^2 - \eta'\omega^4$ will also rotate the contours. By transforming the $\{(1,0),(0,1)\}$ basis vectors to the $\{(1,\eta'),(\eta,1)\}$ basis vectors in the (φ_2, φ_4) space with $\eta_0 = 0.005$ and $\eta'_0 = 500$, it is possible to achieve non-twisted contours with principal axes parallel to the control axes, which have comparable slopes along both new axes in the vicinity of the origin; i.e., for normalized SHG energies of greater than ~82%.

Many fourth order orthogonal polynomial functions, such as Legendre functions, are proportional to $\omega^4 - \eta\omega^2$ by discarding their constant terms. So, parameterization with an orthogonal basis function can to some extent facilitate and accelerate the diagonalization done

by the CMA-ES, while the main job of orthogonalizing parameters is still done by the algorithm. It is also seen in Fig. 7(a) that the initial rate of learning by Legendre parameterization is slightly more than that of the polynomial parameterization.
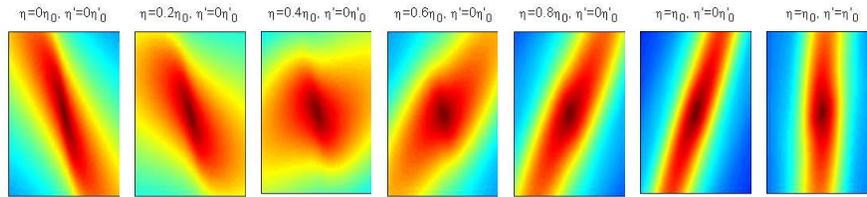

Fig. 10. The modified contour plots of the total SHG energy parameterized with the second order and the fourth order modified polynomial functions. The parameters η and η' are the coefficients of the second/fourth order polynomial added to a fourth/second order polynomial function of the normalized frequency. The horizontal/vertical axes represent the fourth/second order phase terms, respectively. The ranges of coefficients and the color maps are the same in all contour plots. Dark blue in the color map represents zero, and dark red represents one.

5.2 Noise

Figure 7 shows that the problem of adaptive SHG optimization is another case of adaptive control, in which orthogonal parameterization improves the robustness to noise [26,27].

   The noise sources used in our simulations approached the experimental conditions well enough to show reasonable behaviour and to predict the patterns observed experimentally. Alternative sources or parameters can model other physical sources of noise. For example, a unity-amplitude complex noise can be multiplied by the electric filed of a laser pulse to model a negligible shot-to-shot variance of pulse energy, but a relatively unstable phase; i.e., a stable single-shot spectrum and an unstable single-shot SHG spectrum.

   An interesting feature of orthogonal basis functions is their capability of identifying a local optimum point and re-exploring the landscape for another optimum point (see Fig. 7). This behavior is not noticeable in the presence of noise. Note that the fitness value of 0.8 achieved in the presence of noise is close to the fitness value at the first local minimum found by these two parameterizations in noise-free simulations. Small changes in the fitness around a local optimum lead the algorithm to another optimum point with a higher fitness. However, they can be overshadowed by noise, and the algorithm will be trapped in the local optimum point.

   It is important to generate different noise signals and also different distributions for CMA calculations in each run to consider the stochastic nature of the optimization and to have fair comparisons of results. It requires proper initialization and later use of the seed or the state of pseudorandom number generators. A mathematical software package may always use the same state when launched, despite using different states after successive calls (the default setting of MATLAB).

5.3 Frequency-resolved feedback and goal spectrum

A frequency-resolved signal will provide the algorithm with more information about the experiment. The positive impact of this additional information on the learning process can be considerably more than the negative impact of the loss introduced by the spectrometer. For example, a better estimation of the SHG signal is possible by normalizing the measured spectrum to the phase-matching spectrum of the crystal or a goal spectrum as defined below.

   Variations of the phase signal are usually larger at the tails of the SHG spectrum and far from the center wavelength. But the intensity is lower and the signal and noise levels are closer at these wavelengths. A small improvement in the intensity of the SHG in this region is very important, but it has a small contribution to the overall fitness (total energy). Giving more weights to the tails of the SHG spectrum in calculating the fitness can result in a better learning. An empirical rule of thumb is to choose a Gaussian goal spectrum centered at the center of the SHG signal with a larger (almost twice) bandwidth. Our experimental results

show a positive impact on the learning capability of the algorithm using frequency-resolved detection and goal spectrum, even in the case of highly structured SHG profiles.

This improved performance by enhancing the contribution of the tail of the spectrum can be compared with the improved numerical accuracy and stability of a polynomial fit, when the data points are sampled non-uniformly and concentrate at the tails of the function with a Chebyshev distribution. This viewpoint may also be useful for intuitive assessment of the performance of an orthogonal basis function based on its weight function. Note that while the weight function of Legendre functions is a constant, Hermite and Chebyshev functions of the second kind have weight functions that vanish at the two ends of their range of definition.

5.4 Applicability and generality of the results

The results reported here will benefit any experiment in pulse shaping spectroscopy, since all such experiments need to start from a transform-limited pulse at the sample position. Also, a successful optimization of SHG implies successful optimization of non-resonant two-photon absorptions, as in two-photon fluorescence microscopy. It is also possible to use the same setup and optimization algorithm to solve adaptive pulse shaping problems with arbitrary solutions, simply by programming the pulse shaper with an arbitrary spectral phase profile minus the background phase.

Another point is whether successful orthogonal parameterization of CMA-ES for SHG optimization can also be successful for the main optimization in a shaping spectroscopy experiment with another sample, another feedback signal, and a different landscape. The strict answer can only be found after an extensive and diverse set of experiments. However, we anticipate a conditionally affirmative answer: *Orthogonal parameterization of ES is expected to be also robust and efficient in another pulse shaping experiment, if the basis function is chosen with intuition about the nature of the experiment and the form of the fitness landscape.* For example, re-exciting an excited molecule with a second pulse during the dephasing or the relaxation time of the first excited state is a common control strategy. However, the number of re-excitations and the relative amplitudes and phases of subsequent pulses depend on the specific problem. A simple and efficient way of generating such pulse trains in the time domain is to use a single sinusoidal phase function in the frequency domain. Using this simple parameterization, a significant increase of the fitness has been demonstrated, even in the complex biological process of photosynthesis [6] or in an artificial light harvesting complex [7]. The amplitudes and phases of such pulses are Bessel function harmonics in the time domain. This constraint of equally spaced pulses can be relaxed by using a sinusoidal basis function or a Fourier series or in general a sum of (possibly anharmonic) sinusoidal functions.

Binary orthogonal functions have also been used for pulse shaping [23]. However, they generate only a specific class of phase profiles useful in optical telecommunication and some resonant spectroscopy experiments. They are especially useful, when the phase profile is flat over some interval. Modeling a flat profile with polynomials requires higher orders that would result in large variations far from the flat region.

An optimal number of 192 (found from simulations) was used for the number of parameters in the case of direct parameterization. Simulations show that decreasing the parameter size down to 100 reduces the final fitness by 8%. Given the small sensitivity of the final fitness to the parameter size in this range, no significant change is expected by changing the interpolation method. However, for nearly-transform-limited pulses with a small range of variation of phase and no phase wrapping (as opposed to the phase profile we see in the lab and used in our simulations), a smaller optimal parameter size and a faster learning rate is expected. While the comparison between direct and other parameterizations has been fair in our case, a better performance of direct parameterization is possible in other cases, especially if used with spline or more smooth interpolating profiles [15,28].

5.5 Nontrivial benchmark for convergence

Depending on the phase of the input light and the employed parameterization, the landscape of the SHG optimization problem for a given spectrum is reduced to one of its subsets. As

shown in Fig. 3, phase profiles with only the second- and the third-order phase terms have relatively smooth landscapes. However, the presence of the fourth (and higher) order phase terms, as in our simulations, results in a more structured or twisted landscape.

With the same input phase profile, the presence of noise can overshadow small improvements of the fitness function and stop further exploration of the landscape. So, the performances of two different parameterizations cannot be compared with different types or levels of noise. Also note that the "faster convergence" of most parameterizations in the presence of noise in Fig. 7b is associated with a reduction in the final fitness.

In experimental studies, different pulse shaper setups can differ in the resolution/range of shaping, linearity, leakage, and other parameters. Also, different experiments approximate the conditions of Section 2.3 to different extents. The spatiotemporal distortions introduced by the focusing lens of the SHG setup are more pronounced, if the beam incident on the lens already has a structured spatial profile or spatial chirp. These distortions are usually more important in the case of amplified laser pulses that also suffer from more noise and are more commonly used in spectroscopy. Different levels of these distortions make the comparison of experimental results from two parameterizations, one obtained with a femtosecond oscillator and the other with a femtosecond amplifier, even more difficult. A good experimental result with a femtosecond oscillator may or may not be repeated with an amplified laser beam in a comparable configuration.

A fair comparison of different parameterizations requires the disentanglement of all above-mentioned parameters. With that said, it is still insightful to see our results in the light of the results reported in the literature. The convergence of Legendre parameterization with a population size of 40 in 35/31 generations in our simulations/experiments is equivalent to ~1400/1240 individual optimizations, or trials. The minimum number of trials required in experimental results of a recent study [15] is ~1200 (estimated from Fig. 6a of Ref. 15), which is close to the minimum number of trials needed in our experiments.

6. Conclusions

We have demonstrated parameterization schemes for fast and robust optimization by advanced evolution strategy for adaptive laser pulse shaping. Experimental results show a four fold increase in the rate of convergence and ten percent increase in the final fitness of the optimization by replacing the conventional parameterization by one employing Legendre polynomials. Experimental results agree with those obtained from simulations, in which four types of noise model detection fluctuations and laser instability. A detailed analysis of the fitness landscape with 1D, 2D, and 3D scans has been presented to get better insight into the optimization problem and its correlation with the shape and topology of the fitness landscape. The findings of this research benefit any experiment in pulse shaping spectroscopy.