

The Second Harmonic Generation Case-Study as a Gateway for ES to Quantum Control Problems

Ofer M. Shir

Natural Computing Group, Leiden University
Niels Bohrweg 1, Leiden, The Netherlands

oshir@liacs.nl

Thomas Bäck*

Natural Computing Group, Leiden University
Niels Bohrweg 1, Leiden, The Netherlands

baeck@liacs.nl

ABSTRACT

The *Second Harmonic Generation* (SHG), a process that turns out to be a good test case in the physics lab, can also be considered as a fairly simple theoretical test function for global optimization. Despite its symmetry properties, that will be derived here analytically, it seems to capture the complexity of the Fourier transform between the decision space to the evaluation space, and by that to challenge optimization routines. And indeed, counter-intuitively to some extent, locating its global maximum seems to be not an easy task for Evolutionary Algorithms (EAs).

Although this research originates from the real-world applications domain, it aims to introduce a theoretical test case to Evolution Strategies (ES), being a possible theoretical gateway to the real-world physics regime of quantum control problems. After presenting some theoretical results, this paper introduces the study of the *scalability* of the decision space subject to optimization by specific variants of Derandomized Evolution Strategies. We show that the Evolution Strategy in use requires a quasi-quadratic increase of function evaluations for locating the global maximum as the dimensionality increases.

Categories and Subject Descriptors

I.2.8 [Computing Methodologies]: ARTIFICIAL INTELLIGENCE—*Problem Solving, Control Methods, and Search*

General Terms

Algorithms, Experimentation, Performance

Keywords

Second Harmonic Generation, Derandomized Evolution Strategies, Scalability, Quantum Control, Fourier Analysis

*NuTech Solutions, Martin-Schmeisser-Weg 15, 44227 Dortmund, Germany.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO '07, July 7–11, 2007, London, England, United Kingdom.
Copyright 2007 ACM 978-1-59593-697-4/07/0007 ...\$5.00.

1. INTRODUCTION

Evolutionary Algorithms (EAs) are a set of general purpose probabilistic search methods, which are based upon the theory of natural evolution. EAs have three main streams [1]: *Genetic Algorithms* (GAs), developed by J. Holland in the U.S., *Evolution Strategies* (ES), developed in Germany by I. Rechenberg and H.P. Schwefel, and *Evolutionary Programming* (EP), developed by L.J. Fogel et al. in the U.S. Evolution Strategies are a *canonical EA for continuous function optimization*, due to their straightforward continuous encoding, their specific variation operators, as well as to their high performance in this domain in comparison to other methods on benchmark problems. The higher the dimensionality of the search space, the more suitable a task becomes for an ES (see, e.g. [1]).

This particular study originates in the application of ES to the optimization of numerical *Quantum Control* physics problems, where the so-called Second Harmonic Generation (SHG) was suggested as a theoretical test-case for the algorithms in use, as a “warming-up” before the primary optimization task, which was time-consuming. It was shown experimentally that the performance of Evolution Strategies on the two problems was highly correlated, and thus provided with motivation to investigate the SHG.

Although this function has some mathematical properties that could suggest an easy search for its global maximum, it is yet a difficult task for a global optimizer in general, and for an EA in particular. Some of its difficulty seems to lie within the complexity of the Fourier transform between the decision space to the evaluation space. We consider the SHG as an interesting theoretical test problem, with a strong link to real-world physics problems, that introduces a type of complexity which has not been studied before in the context of ES.

Here we present formally the Second Harmonic Generation optimization problem, and use it as a case-study for investigating the *scalability* of Evolution Strategies, i.e. the performance of the optimization routine as a function of the problem’s dimensionality. The reader should note that in the SHG context, the scalability is applied to the decision space, but does not affect the objective landscape - as will be described here. Hence, this scalability study is not equivalent to a general dimensionality scan of a test-problem.

We would like to mention that a study of the control of multiphoton transitions by means of shaped laser pulses was presented recently [4], where SHG-like processes were discussed, and evolutionary optimization approach (GA-based)

was applied for learning phase profiles. This paper of ours may suggest an extension, to some limited degree, of that study.

The remainder of the paper is organized as follows. Section 2 provides the reader with the motivation for this study, the Quantum Control field. Section 3 presents the Second Harmonic Generation and its mathematical properties. In section 4 we introduce the evolutionary algorithms in use - the family of Derandomized Evolution Strategies. This is followed in section 5 by the description of the experimental setup, a preliminary test and the numerical final results. In section 6 we draw conclusions, summarize our study, and propose future directions in the domain of our study.

2. THE MOTIVATION: QUANTUM CONTROL PROBLEMS

The advent of modern laser pulse shaping techniques in the femtosecond regime has made it possible to control the motion of nuclei and even electrons by a judicious choice of the pulse shapes. The application to dynamic molecular alignment [7] is of considerable interest in this context because of its many practical consequences. There is currently a great interest in the atomic and molecular physics community to align molecules with laser pulses, since dealing with an aligned sample of molecules simplifies the interpretation of experimental data: a multitude of chemical and physical processes ranging from bimolecular reactions to high harmonic generation are influenced by the angular distribution of the molecular sample. Furthermore, in many fundamental molecular dissociation or ionization experiments, the interpretation of the collected data becomes much easier when the molecules are known to be aligned with respect to a certain axis. Hence, techniques to generate molecular alignment are much needed.

The goal of this research field is thus to optimize the *alignment* of an ensemble of molecules after the interaction with a shaped laser pulse. By applying a self-learning loop using an evolutionary mechanism, the interaction between the system under study and the laser field can be steered, and optimal pulse shapes for a given optimization target can be found. So far, the role of the experimental feedback in the self-learning loop has been mainly played by a *numerical simulation*.

Dynamic Molecular Alignment

To calculate the time-dependent alignment, the Schrödinger's equation for the angular degrees of freedom of a model diatomic molecule under the influence of the shaped laser field is solved. Explicitly, the time-dependent profile of the pulse, which completely determines the dynamics after the transition to the rotating frame has been performed, is described by:

$$E(t) = \int_{-\infty}^{\infty} A(\omega) \exp(i\phi(\omega)) \exp(i\omega t) d\omega, \quad (1)$$

where $A(\omega)$ is a Gaussian window function describing the contribution of different frequencies to the pulse and $\phi(\omega)$, the *phase function*, equips these frequencies, which are equally distributed across the spectrum of the pulse, with different complex phases. Hence, by changing $\phi(\omega)$, the temporal structure of $E(t)$ can be altered.

The careful reader should note that $E(t)$ cannot be written in a closed-form, in most of the cases, since the analytical solution of a Fourier-transformed arbitrary complex function is usually impossible to obtain.

In a real life pulse shaping experiment, $A(\omega)$ is fixed and $\phi(\omega)$ is used to control the shape of the pulses. The same approach is usually used in numerical simulations, i.e. **the search space is in the frequency domain while the function evaluation is performed in the time domain**. $\phi(\omega)$ is interpolated at n frequencies $\{\omega_i\}_{i=1}^n$; the n values $\{\phi(\omega_i)\}_{i=1}^n$ are the decision parameters to be optimized. After the evaluation of the field-molecule interaction, the alignment's quantity is defined as the expectation value of the *cosine-squared* of the angle of the molecular axis with respect to the laser polarization axis. However, the description of the molecule as a rigid rotator is strictly valid only for low field intensities, and the higher the applied intensity, the more important other competing channels like dissociation and ionization will become. Therefore, one would like to achieve good alignment while keeping the peak laser intensity as low as possible.

This quantum control problem has been tackled at several levels. A recent study presented a survey of modern evolutionary approaches to the problem, and showed that it payed off to use more elaborated optimization schemes, and in particular Derandomized Evolution Strategies, for such a high-dimensional optimization task [6].

3. SECOND HARMONIC GENERATION

A measure of field intensity is given by the *Second Harmonic Generation*, as will be described here. A laser pulse going through certain crystals produces light at the octave of its frequency spectrum. The total energy of the radiated light is proportional to the integrated squared-intensity of the primary pulse. The time-dependent profile of the laser field is exactly as given in Eq. 1.

The *Second Harmonic Generation* signal is then defined by:

$$SHG = \int_0^{\infty} |E(t)|^4 dt \quad (2)$$

i.e., an integration over time of the intensity. Since $E(t)$ is usually cannot be given in a closed-form, this integral has to be calculated numerically. Second Harmonic Generation (SHG) is a process that turns out to be a good test case in the physics lab, and its investigation contributes to the understanding of the alignment problem. This is due to the fact that the SHG is a measure of the spikiness of a pulse, and this property is useful for the definition of a punishment function for the dynamic molecular objective function, due to the sole validity of low field intensities. Moreover, a multi-criteria approach has been applied to the alignment problem, where the SHG is playing a key role as one of the criteria. The cosine-squared alignment is the primary objective, subject to maximization, and the minimization of the SHG was set as the secondary objective, for the fulfillment of the rigid-rotor physics approximation. Preliminary results of that study reveal the nature of the trade-off between the angular alignment and the intensity, which was expressed through the *second harmonic generation*, and it seems that the importance of the intensity criterion is likely to govern the decision of the expert on the trade-off surface, which is to look for solutions with relatively good cosine-squared alignment values in the region of fair trade-offs.

From the theoretical point of view, the SHG is a simple function, with some interesting properties that will be shown here, but yet not an easy optimization task for global optimizers.

3.1 Mathematical Properties

The following section is mainly based on Bracewell [2]. In order to show some properties of this function, it is convenient to work in the frequency space.

Definition 1. Given the spectral amplitude equipped with the complex phases, $E(\omega) = A(\omega) \exp(i\phi(\omega))$, consider its autocorrelation (convolution) function $\tilde{E}(\omega)$:

$$\begin{aligned} \tilde{E}(\omega) &= E(\omega) * E^*(\omega) = \\ &= \int_{-\infty}^{\infty} E^*(\Omega) \cdot E(\Omega + \omega) d\Omega = \int_{-\infty}^{\infty} E(\Omega) \cdot E^*(\Omega - \omega) d\Omega \end{aligned}$$

where $E^*(\omega)$ denotes the complex conjugate.

We would like to show how this autocorrelation function in the frequency domain is linked to the time domain:

THEOREM 1. *The autocorrelation function of the spectral amplitude, $\tilde{E}(\omega)$, is equal to the Fourier transform of the time-dependent intensity function, i.e.:*

$$\tilde{E}(\omega) = \int_{-\infty}^{\infty} |E(t)|^2 \exp(-i\omega t) dt \quad (3)$$

PROOF.

$$\begin{aligned} \tilde{E}(\omega) &= \int_{-\infty}^{\infty} E^*(\Omega) \cdot E(\Omega + \omega) d\Omega = \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} E^*(t) \exp(i\Omega t) dt \right] \cdot \\ &\quad \left[\int_{-\infty}^{\infty} E(\tau) \exp(-i(\omega + \Omega)\tau) d\tau \right] d\Omega = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E^*(t) E(\tau) \exp(-i\Omega(t - \tau)) \cdot \\ &\quad \exp(-i\omega\tau) d\Omega dt d\tau = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E^*(t) E(\tau) \delta(\tau - t) \exp(-i\omega\tau) dt d\tau = \\ &= \int_{-\infty}^{\infty} E^*(t) E(t) \exp(-i\omega t) dt = \int_{-\infty}^{\infty} |E(t)|^2 \exp(-i\omega t) dt \end{aligned}$$

where $\delta(x - \tilde{x})$ is the Dirac delta function. \square

THEOREM 2. (Plancherel's Theorem) *Given $f(x)$, which has the Fourier transform $F(s)$, the integral over the squared modulus of $f(x)$ is equal to the integral over the squared modulus of its spectrum $F(s)$:*

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \int_{-\infty}^{\infty} |F(s)|^2 ds$$

See Bracewell [2] for the proof.

Thus, we can conclude from theorems 1 and 2 that

$$\int_{-\infty}^{\infty} |\tilde{E}(\omega)|^2 d\omega = \int_{-\infty}^{\infty} |E(t)|^4 dt = SHG \quad (4)$$

Global Maximum

THEOREM 3. *The SHG is maximized by a zero phase function:*

$$\operatorname{argmax}_{\phi(\omega)} \{SHG(\phi(\omega))\} \equiv 0$$

PROOF. It is convenient to show this in the frequency domain. Using the result obtained at Eq. 4, we may write

$$\begin{aligned} SHG &= \int_{-\infty}^{\infty} |\tilde{E}(\omega)|^2 d\omega = \\ &= \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} E(\Omega) \cdot E^*(\Omega - \omega) d\Omega \right|^2 d\omega = \end{aligned}$$

Apply a change of variables $\tilde{\omega} = \Omega - \frac{\omega}{2}$, and **consider only the integrand**:

$$\begin{aligned} &\left| \int_{-\infty}^{\infty} E\left(\tilde{\omega} + \frac{\omega}{2}\right) \cdot E^*\left(\tilde{\omega} - \frac{\omega}{2}\right) d\tilde{\omega} \right|^2 = \\ &= \left| \int_{-\infty}^{\infty} A\left(\tilde{\omega} + \frac{\omega}{2}\right) \cdot A\left(\tilde{\omega} - \frac{\omega}{2}\right) \cdot \right. \\ &\quad \left. \exp\left\{i\left[\phi\left(\tilde{\omega} + \frac{\omega}{2}\right) - \phi\left(\tilde{\omega} - \frac{\omega}{2}\right)\right]\right\} d\tilde{\omega} \right|^2 \end{aligned} \quad (5)$$

and the integral is now split into real functions of the spectrum, $A(\omega)$, and the appropriate complex phases. From elementary theory of complex numbers, we know that a sum over complex numbers can only be maximized if all its elements share the same phase, otherwise a disruptive superposition effect would occur. Here, due to the symmetric nature of the phase term in the integrand, $[\phi(\tilde{\omega} + \frac{\omega}{2}) - \phi(\tilde{\omega} - \frac{\omega}{2})]$, which has to occur for every ω , it is obvious that only a constant term could generate equal phases, i.e.

$$\forall \omega \quad \phi\left(\tilde{\omega} + \frac{\omega}{2}\right) = \phi\left(\tilde{\omega} - \frac{\omega}{2}\right)$$

and thus only a constant (flat) phase can maximize the integral. Among all the constant functions, the zero phase function is the global maximum of the SHG, due to an integration constant. \square

Symmetry

THEOREM 4. *Mirror-image symmetry with respect to every axis is required by the phase function to enhance the SHG, i.e., given a phase function, an increase in the SHG value can be achieved only if each phase point is equal to its mirror-image-point with respect to every mirror-axis.*

PROOF. Rewriting Eq. 5 in terms of "left" (L) and "right" (R) elements in the integrand (with respect to $\frac{\omega}{2}$), one obtains:

$$SHG = \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} A(R) \cdot A(L) \cdot \exp\{i[\phi(R) - \phi(L)]\} d\tilde{\omega} \right|^2 d\omega$$

Using the same argumentation given in the previous proof, it is clear that only mirror-image symmetry with respect to $\frac{\omega}{2}$, **for every value of ω** , could enhance the SHG value, and thus the theorem holds. \square

Theorem 4 originates from the *convolution* nature of the construction process. Figure 1 provides the reader with an illustration for the so-called mirror-image effect - the contribution of two phase points around the central frequency ω_0

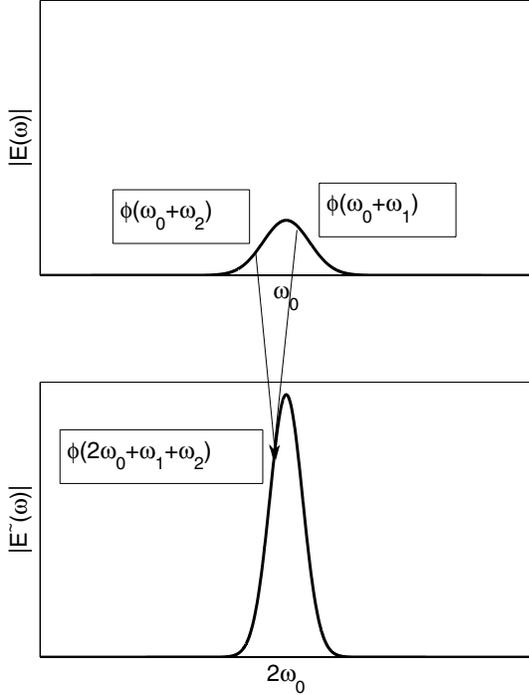


Figure 1: Construction of $\tilde{E}(\omega)$ out of $E(\omega)$.

at $E(\omega)$, $\phi(\omega_0 + \omega_1)$ and $\phi(\omega_0 + \omega_2)$, to the construction of $\tilde{E}(\omega)$ with $\phi(2 \cdot \omega_0 + \omega_1 + \omega_2)$. Notice the shift in the central frequency, and the scaling of the Gaussian.

3.2 The SHG Optimization Problem

We hereby formally present the SHG test problem. Introduce the *unnormalized discrete form* of the SHG,

$$f_{SHG}(\vec{\varphi}) = \sum_{k=0}^{n-1} |E_k|^4 = \sum_{k=0}^{n-1} \left| \sum_{l=0}^{n-1} A_l \cdot \exp\{i\varphi_l\} \exp\left\{\frac{2\pi i}{n} kl\right\} \right|^4 \quad (6)$$

and consider its *maximization*:

$$f_{SHG}(\vec{\varphi}) \rightarrow \max$$

Maximizing f_{SHG} is equivalent to finding a vector $\vec{\varphi}^* \in \mathbb{R}^n$ which satisfies

$$\forall \vec{\varphi} \in \mathbb{R}^n : f_{SHG}(\vec{\varphi}) \leq f_{SHG}(\vec{\varphi}^*) \equiv f_{SHG}^*$$

Based on the theorems given in the previous section, we know that the SHG maximization problem can be solved by locating a flat (zero) phase function.

Intuitively, this seems to be an easy optimization task, maybe even a continuous equivalent problem to the *counting ones* (zeros) problem for GAs [1], with a trivial global maximum. On top of that, the symmetry property looks as if it could assist the optimizer to find the global maximum in a fairly easy way. On the other hand, the search space is in the frequency domain while the fitness evaluation is performed in the time domain (or after *convolution*,

respectively), and this transformation seems to capture the difficulty for an optimizer.

One can draw an analogy between the *frequency-time transformation* and the *genotype-phenotype mapping* of the traditional GA [1]. In essence, the decision parameters to be optimized, the analogue to the *genotype*, parameterize the phase function in the frequency domain. The Fourier transform then maps the phase function onto the time-dependent electric field. The latter undergoes an evaluation of its fitness, *integration* in our case, and thus plays the analogue role of the phenotype.

As for the numerical details, to this end $\phi(\omega)$ has been interpolated at n frequencies $\{\omega_i\}_{i=1}^n$; the n values $\{\phi(\omega_i)\}_{i=1}^n$ are our decision parameters to be optimized.

The global maximum is normalized to 1:

$$\max \{f_{SHG}\} = f_{SHG}(\vec{0}) = 1 \quad (7)$$

Yet, it had been already shown experimentally that it was not an easy task at all for an evolutionary algorithm to find the global maximum of the SHG when parameterizing the phase function with at least $n = 80$ function values to be interpolated.

A survey of EAs, which was applied to the *dynamic molecular alignment* optimization task, considered the SHG maximization as a preliminary task for the algorithms in use [6]. Most of the techniques, ranging from the *traditional GA* and the *standard ES* to sophisticated derandomized ES variants, failed to find a good solution to the SHG maximization problem after 10000 function evaluations. Note that for the binary-coded GA that was applied for this task, this problem is defined as the maximization of the *counting zeros* problem - and yet it performed poorly.

Only two derandomized ES variants performed well and were clearly superior with respect to the other techniques. Moreover, that study showed a **consistency between the performance per algorithm on the SHG versus the alignment problem, which provides us with further motivation to investigate the SHG.**

3.3 Problem Difficulty: Numerical Assessment

In order to assess the complexity of the SHG maximization problem, and as a numerical validation for the theoretical results presented in the previous section regarding the symmetry property of the function, we conducted the following simple statistical test. We consider a phase function with $n = 100$ function values, which are randomly initialized in the interval $[0, 2\pi]$, and three different routines to maximize the SHG: (1) setting function values to zero when consistently indexing from right to left, (2) consistent indexing from left to right, and (3) random permutation of indices, with no repetition, setting values to zero. These routines were run 100 times, while the SHG values were recorded at each time-step per routine. Figure 2 presents the mean values and standard deviations per time-step for the three routines. It is clear from this plot that around 45% of the function values must be set to 0 in the consistent indexing in order to enhance the SHG value, no matter what was the direction of the scan. This is expected from theory - due to the shape of the Gaussian, there is only a negligible contribution to the SHG value from points which are far from the central peak, and only around that frequency the value dramatically increases, subject to the “mirror-image” condition (theorem 4). The random permutation routine can set

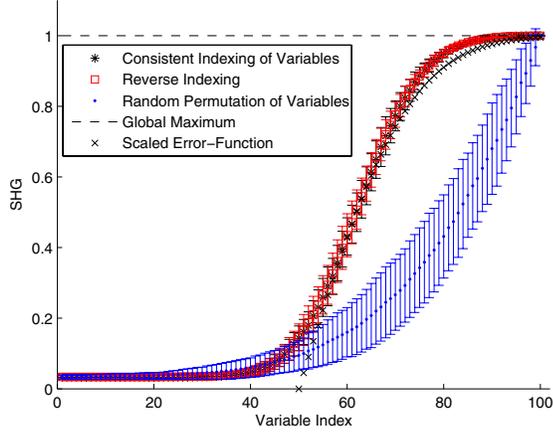


Figure 2: Setting function values to 0: numerical results over 100 runs.

two mirror-image phase points to zero at every time-step, even in the area of the central frequency, and hence it has a different profile of mean values and their standard deviation respectively. Still, on average, its SHG value does not “take-off” until approximately $\frac{1}{3}$ of the function values are set to 0. It is interesting to observe how close the profile of the SHG of the consistent indexing (routines (1) and (2), direct as well as reverse) is to the positive half of a scaled *error function*,

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

This also suggests how slow the descent from the SHG global maximum value towards sub-optimal solutions is.

4. ALGORITHMS

Based on previous experience with quantum control problems, and due to experimental results that showed that certain variants of Derandomized Evolution Strategies are superior with respect to other Evolutionary Algorithms on those problems, we restrict our study to these algorithms. Here, we provide a short background of the specific variants in use.

4.1 Derandomized Evolution Strategies

Mutative step-size control tends to work well for the adaptation of a global step-size, but tends to fail when it comes to the individual step-size. This is due to several disruptive effects [3] as well as to the fact that the selection of the *strategy parameters* setting is indirect, i.e. not the vector of a successful mutation is used to adapt the step-size parameters, but the parameters of the distribution that led to this mutation vector. The so-called *derandomized mutative step-size control* aims to tackle those disruptive effects. The first versions of *derandomized ES algorithms* introduced a controlled global step-size in order to monitor the individual step-sizes by decreasing the stochastic effects of a probabilistic sampling. The selection disturbance was completely removed with later versions by omitting the adaptation of strategy parameters by means of probabilistic sampling. This was combined with individual information from the last gener-

ation (the successful mutations, i.e., of selected offspring), and then adjusted to *correlated mutations*. Later on, the concept of *adaptation by accumulated information* was introduced, aiming to use wisely the past information for the purpose of step-size adaptation: instead of using the information from the last generation only, it was successfully generalized to a weighted average of the previous generations.

It is important to note that the different variants of *derandomized ES* hold different numbers of strategy parameters to be adapted, and this is a factor in the learning speed of the optimization routine: it is either a linear or quadratic order in terms of the dimensionality of the search problem n , and there seems to be a trade-off between the number of strategy parameters and the time needed for the adaptation/learning process of the step-sizes.

An explicit description follows.

The $(1, \lambda)$ -DR2 Algorithm

The **DR2** Algorithm [5] is considered to be the second generation of the derandomized Evolution Strategies. This variant uses a linear number in n of strategy parameters, and it aims to accumulate information about the correlation or anti-correlation of past mutation vectors in order to adapt the step-size:

$$\vec{x}^{g+1} = \vec{x}^g + \delta^g \bar{\delta}_{scal}^g \vec{Z}^k \quad \vec{Z}^k = \vec{\mathcal{N}}(0, 1) \quad (8)$$

$$\vec{Z}^g = c \vec{Z}_{sel} + (1 - c) \vec{Z}^{g-1} \quad (9)$$

$$\delta^{g+1} = \delta^g \cdot \left(\exp \left(\frac{|\vec{Z}^g|}{\sqrt{n} \sqrt{\frac{c}{2-c}}} - 1 + \frac{1}{5n} \right) \right)^\beta \quad (10)$$

$$\bar{\delta}_{scal}^{g+1} = \bar{\delta}_{scal}^g \cdot \left(\frac{|\vec{Z}^g|}{\sqrt{\frac{c}{2-c}}} + 0.35 \right)^{\beta_{scal}} \quad (11)$$

where $\vec{\xi}_{scal} = \vec{\mathcal{N}}(0, 1)^+$, $\vec{Z} \in \{-1, +1\}^n$, and β , β_{scal} , b and ξ^k are constants.

The (μ_W, λ) Covariance Matrix Adaptation ES

The (μ_W, λ) -**CMA-ES** algorithm [3] is known as the state-of-the-art among of the derandomized ES variants (could also be considered as DR4). It has been successful for treating correlations among object variables, where it applies *principal component analysis* (PCA) to the *selected mutations* during the evolution, also referred to as “*the evolution path*”, for the adaptation of the covariance matrix of the distribution. The concept of *weighted recombination* is introduced: applying intermediate multi-recombination on the best μ out of λ with given weights $\{w_i\}_{i=1}^\mu$. The result is denoted with $\langle \vec{x} \rangle_W$. Furthermore, $\vec{p}_c^{(g)} \in \mathbb{R}^n$ is the so-called *evolution path*, the crucial component for the adaptation of the covariance matrix, and $\vec{p}_\sigma^{(g)} \in \mathbb{R}^n$ is the *conjugate evolution path*, which is responsible for the step-size control. $\mathbf{C}^{(g)} \in \mathbb{R}^{n \times n}$, the covariance matrix of the mutation distribution ($\mathbf{C}^{(g)} = \mathbf{B}^{(g)} \mathbf{D}^{(g)} (\mathbf{B}^{(g)} \mathbf{D}^{(g)})^T$):

$$\vec{x}^{g+1} = \langle \vec{x} \rangle_W + \sigma_g \mathbf{B}^g \mathbf{D}^g \vec{z}_k^{g+1} \quad (12)$$

$$\vec{p}_c^{g+1} = (1 - c_c) \cdot \vec{p}_c^g + c_c^u \cdot c_W \mathbf{B}^g \mathbf{D}^g \langle \vec{z} \rangle_W^{g+1} \quad (13)$$

$$\mathbf{C}^{g+1} = (1 - c_{cov}) \cdot \mathbf{C}^g + c_{cov} \cdot \bar{p}_c^{g+1} (\bar{p}_c^{g+1})^T \quad (14)$$

$$\bar{p}_\sigma^{g+1} = (1 - c_\sigma) \cdot \bar{p}_\sigma^g + c_\sigma^u \cdot c_W \mathbf{B}^g \langle \bar{z} \rangle_W^{g+1} \quad (15)$$

$$\sigma^{g+1} = \sigma^g \cdot \exp\left(\frac{1}{d_\sigma} \cdot \frac{\|\bar{p}_\sigma^{g+1} - \hat{\chi}_n\|}{\hat{\chi}_n}\right) \quad (16)$$

where $\hat{\chi}_n$ is the expected length of \bar{p}_σ .

c_c , c_{cov} , c_σ and d_σ are learning/adaptation rates, $\{w_i\}_{i=1}^\mu$ are the recombination weights, and $c_c^u := \sqrt{c_c(2 - c_c)}$, $c_W := \frac{\sum w_i^\mu}{\sqrt{\sum w_i^{2\mu}}}$ and $c_\sigma^u := \sqrt{c_\sigma(2 - c_\sigma)}$ are derived respectively.

All weighting variables and learning rates were applied as suggested in the given citations, and particular in [3].

For more details regarding the *family of derandomized ES* we refer the reader to [3].

5. SCALABILITY: NUMERICAL RESULTS

We hereby present our experimental framework for studying the scalability of Evolution Strategies with respect to the Second Harmonic Generation problem, i.e., the performance of the optimization routine as a function of the problem's dimensionality n .

We begin by testing the natural quality of the various parameterizations. This is followed by the crucial element of this study - selecting the strategy - and then we apply it to the optimization of the function in an increasing dimensions of the problem.

All numerical experiments were executed with **Matlab 7.0** under Suse-Linux.

5.1 Preliminary: Initial States Density Test

The following preliminary experiment is meant to compare the natural initial quality of the different parameterizations. We applied a so-called *initial states density test*, a statistical fitness measurement of the initialized phase functions in the different parameterizations. For each parameterization in use, $n = 80..2000$, we initialized 1000 phase functions and calculated the mean fitness and the standard deviation respectively.

Figure 3 presents the statistical results for this test. The scale of initial values is one order of magnitude: $[0.005, 0.055]$. As n increases, the initial fitness mean value dramatically drops, in particular for $n > 200$. At some point, the initial value stabilizes on $f_{SHG}^{init} \approx 0.005$ with a very low standard deviation.

5.2 Preliminary: Selecting a Strategy

It had been shown in [6] that only two derandomized Evolution Strategies performed well on the SHG maximization task - the **DR2** and the **CMA-ES** algorithms. Due to the low number of experiments (5 runs per algorithm) on which that conclusion was based upon, we conducted a preliminary set of runs in order to select our strategy. The reader should keep in mind that there are dramatic practical consequences for the choice of strategy for this study - the DR2 holds a linear number of strategy parameters in n , whereas the CMA-ES holds a quadratic number of strategy parameters - and since the goal of this research is to study the dimensionality, the CMA-ES is expected to face numerical difficulties far earlier than the DR2.

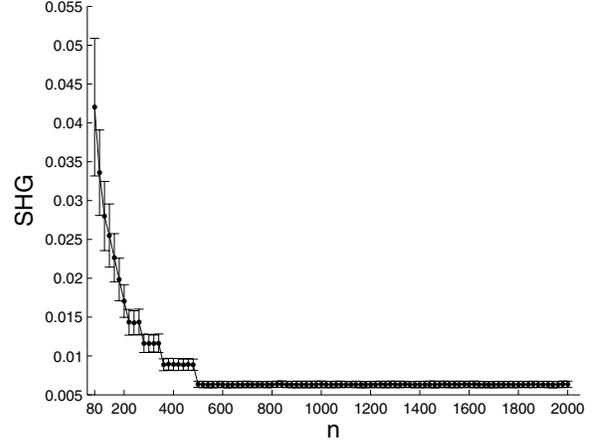


Figure 3: Initial states density test for the various parameterizations, $n = 80..2000$: mean and standard deviation over 1000 random initializations in the interval $[0, 2\pi]$.

Table 1: Maximizing the SHG over 100 runs with $n = 100$ function values: 10^4 function evaluations per run.

SHG	(1, 10)-DR2	(1, 10)-CMA	(8, 17)-CMA
Mean	0.9984	0.9880	0.9629
Std	0.0043	0.0006	0.0822

We considered 100 runs of 10^4 function evaluations of the two strategies in the following settings: (1, 10)-DR2, (1, 10)-CMA-ES and (8, 17)-CMA-ES (the latter is the default population size for the given problem by Hansen et al.). The numerical results of that run are given in table 1. It is clear that the DR2 algorithm outperforms the CMA-ES on this particular problem, and hence it is selected as our strategy for this study.

From the practical point of view this result is positive, for the reason mentioned earlier, and it would allow the investigation of scalability more easily from the numerical-technical perspective.

5.3 Scan over Dimensions

We apply the DR2 algorithm to the SHG maximization problem, with an increasing number of function values. We consider $n_0 = 80$ as the basic parameterization of the phase function, and we go up to $n_{up} = 2000$ function values - with steps of $\Delta n = 10$. We consider two different experiments:

1. Obtaining the maximal SHG value with 10^4 function evaluations; 20 runs per parameterization.
2. Evaluating the required number of function evaluations for reaching the global maximum, with an upper bound of 10^6 evaluations; 10 runs per parameterization.

The result obtained for the first routine is plotted as Figure 4. As can be observed in the plots, the performance in

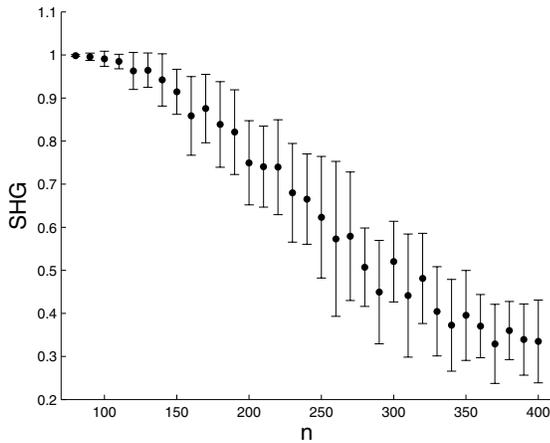
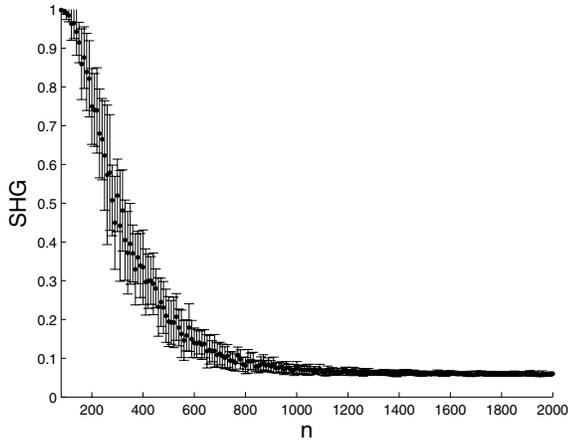


Figure 4: SHG value obtained with DR2 given 10^4 function evaluations; mean and standard deviation over 10 runs per dimension. Top: full scan ($n = 80..2000$); bottom: zoom-in ($n = 80..400$).

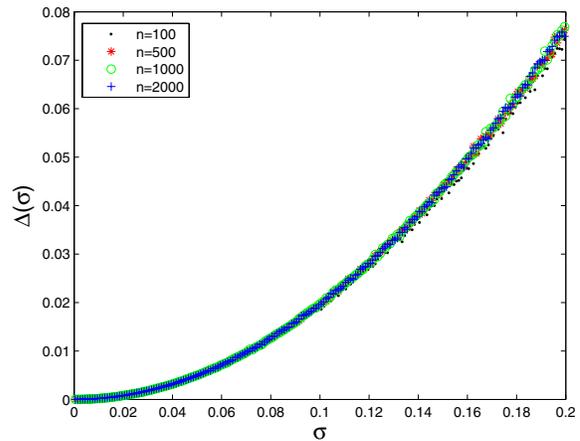
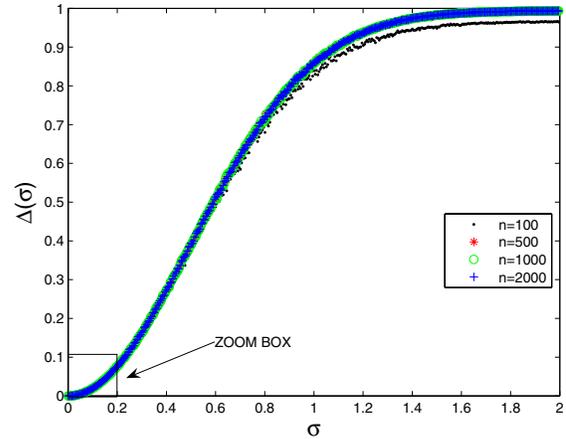


Figure 6: Sensitivity of the global maximum as a function of σ for $n = 100, 500, 1000, 2000$: mean over 100 evaluations per point. Top: full scan ($\sigma \in [0, 2]$); bottom: zoom-in ($\sigma \in [0, 0.2]$).

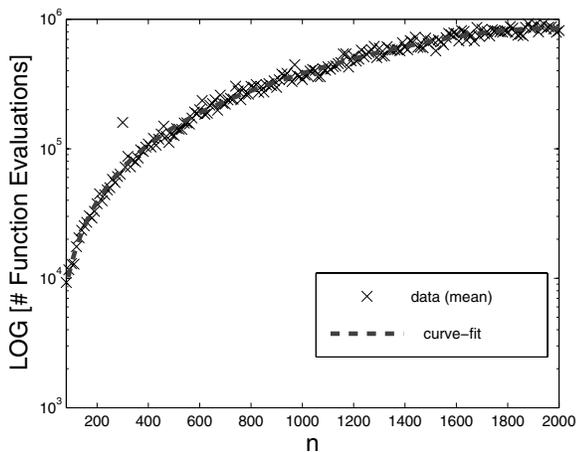


Figure 5: Log number of evaluations required for the DR2 to reach the SHG global maximum, given up to 10^6 function evaluations; mean over 10 runs per dimension, with the curve-fit to the data.

this test-case is dramatically compromised as the dimensionality n increases. For values around $n = 200$ the final result for 10^4 function evaluations is already pretty low, and this trend becomes a complete disaster for $n > 500$. We conclude that the number of function evaluations is a crucial element already for low dimensions. Figure 5 presents the result of the second scanning routine. The number of function evaluations required for the location of the global maximum seems to be a polynomial function of the 3^{rd} degree in the dimensionality n . Explicitly, the curve fitting routine suggested the following function as the curve-fit:

$$\mathcal{F}_{curve-fit} = -0.00063 \cdot n^3 + 0.58 \cdot n^2 + 120 \cdot n + 4800$$

With relatively small coefficient for the 3^{rd} term, one can consider this function as a quasi-quadratic function in the specified domain. This is not an intuitive result, given the first scan-experiment, presented earlier. We can conclude that even though the optimization routine fails to accomplish a successful global maximum location with 10^4 function evaluations as the dimension increases, it requires only a quasi-quadratic increase of function evaluations for this task.

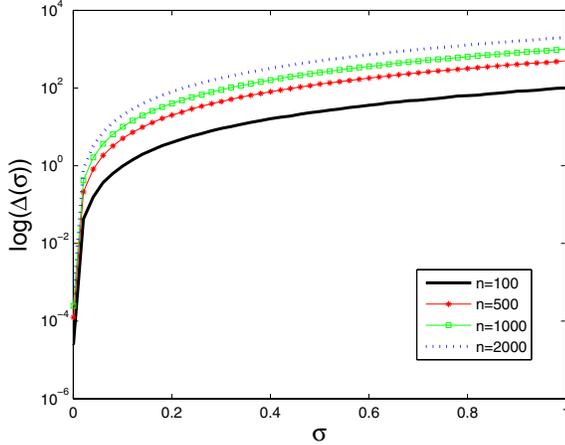


Figure 7: log-sensitivity of the global minimum of the *sphere function* as a function of σ for $n = 100, 500, 1000, 2000$: mean over 100 evaluations per point. Dramatic differences between the dimensions are observed.

5.4 Sensitivity Analysis

Given the numerical results of the previous section, we would like to test the sensitivity of the global maximum in the different parameterizations, i.e., the various scalability dimensions n . We apply the following numerical sensitivity analysis routine: given the global maximum, $\vec{x}^* = \vec{0}$ of dimension n , normally-distributed variations are applied to it with zero-mean and an increasing variance. We then define the sensitivity of the global maximum in dimension n to standard deviation σ as:

$$\begin{aligned} \Delta(n, \sigma) &= \max \{f_{SHG}\} - f_{SHG}(\vec{x}^* + \sigma \cdot \vec{\mathcal{N}}(0, \mathcal{I})) = \\ &= 1 - f_{SHG}(\sigma \cdot \vec{\mathcal{N}}(0, \mathcal{I})) \end{aligned} \quad (17)$$

This routine is applied throughout a scan of σ , whereas $\Delta(n, \sigma)$ is evaluated 100 times per σ and averaged. We applied this test for 4 different parameterizations: $n = 100, 500, 1000, 2000$. The outcome of this test is plotted as Figure 6. This test reveals a *logistic* profile of the sensitivity of the global maximum: an exponentially increasing sensitivity for low values of σ (order of $\sigma = 0.1$), with saturation of 1 for high values (order of $\sigma = 1$). Surprisingly, the cases $n = 500, 1000, 2000$ share the same profile, whereas $n = 100$ seems to have a slightly lower sensitivity for the high values of σ .

We may conclude that the global maximum is robust in the context of our search, in all dimensions, since the order of magnitude of the step-size reaches the regime of $\sigma \approx 0.1$ fairly fast. For comparison, we applied an equivalent sensitivity test to the *sphere function*; see Figure 7. Clearly, the various dimensions have different sensitivities (note the log-scale), but share a parabolic profile.

6. CONCLUSIONS AND OUTLOOK

We have introduced the Second Harmonic Generation as a theoretical test function for global optimization and provided the reader with the motivation for its investigation.

We have derived analytically its mathematical properties, that could have suggested an easy global optimization, but later on showed experimentally that this was not an easy task for the advanced Evolution Strategies. We conclude that the SHG captures the complexity of the Fourier transform between the decision space to the evaluation space, and by that it challenges the optimization routines. We suggested an analogy between the genotype-phenotype mapping and the Fourier transform which acts on the phase function to construct the time-dependent electric field. The Fourier transform is an elementary transformation in physics problems, and in particular in quantum control calibration tasks, as was discussed here. It seems that no existing routine among the Evolutionary Algorithms, except for a specific variant of derandomized ES, the so-called DR2, could tackle the SHG problem in a satisfying manner.

Given that, we have performed a scalability test of this ES variant, with respect to the SHG problem, and concluded that even though the optimization routine fails to accomplish a successful global maximum location with 10^4 function evaluations as the dimension increases, it requires only a quasi-quadratic increase of evaluations for the sake of the global maximum location. Later on we performed an experimental sensitivity analysis of the global maximum, and concluded that it is a robust optimum.

We would like to post a message and encourage the EC community to consider this SHG function as a theoretical test case in future work. Moreover, as our future work we plan to further investigate the scalability of Evolution Strategies with respect to other case-studies. Additionally, we would like to consider the design of a *complex-number* Evolution Strategy, that could be applied directly to the electric field parameters in physics problems.

Acknowledgments

Ofer Shir would like to thank Riccardo Fanciulli and Michael Emmerich for the fruitful discussions. This work is part of the research program of FOM, supported by NWO.

7. REFERENCES

- [1] T. Bäck. *Evolutionary algorithms in theory and practice*. Oxford University Press, New York, NY, USA, 1996.
- [2] R. Bracewell. *The Fourier Transform and Its Applications*. McGraw-Hill Book Company, 1965.
- [3] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.
- [4] V. V. Lozovoy and M. Dantus. Systematic control of nonlinear optical processes using optimally shaped femtosecond pulses. *ChemPhysChem*, 6(10):1970–2000, 2005.
- [5] A. Ostermeier, A. Gawelczyk, and N. Hansen. Step-size adaption based on non-local use of selection information. In *PPSN*, volume 866 of *Lecture Notes in Computer Science*. Springer, 1994.
- [6] C. Siedschlag, O. M. Shir, T. Bäck, and M. J. J. Vrakking. Evolutionary algorithms in the optimization of dynamic molecular alignment. *Optics Communications*, 264:511–518, Aug. 2006.
- [7] H. Stapelfeldt. *Rev. Mod. Phys.*, 75(543), 2003.