Chapter 9

Retrieval of polarization shaped pulses with a multi-objective EA

9.1 Introduction

For pulse forms originating from the parallel setup where a phase shift of one beam-path versus the other has already a great impact on the whole polarization transient, a method to reliably determine the full pulse information is highly desirable. Using conventional, dual-channel interferometry methods like POLLIWOG [160] for pulses originating from the parallel setup would mean placing one interferometer into another interferometer arm, which would be an arduous experimental effort. The alternative, time-resolved-ellipsometry (TRE) [151], requires an additional measurement through a quarter wave plate in order to determine the helicity and, as a pure temporal measurement, does not provide phase information.

A number of cross-correlations to different projection angles already provide a good insight to the temporal course of major axis angle, intensity, and also ellipticity, as shown in the previous chapters. Phase-driven features (like for example chirps) and the helicity can not be retrieved by such measurements. In the previous chapter, spectral intensity projections to different angles (Fig. 8.3, bottom) showed that there is much information contained within the pulse spectra. Recording a set of spectra together with cross-correlations for multiple angles could help assigning helicity and uncover phase information, without resorting to interferometry.

Pooling many spectra and crosscorrelations to one single solution (a polarization shaped pulse) is unfortunately neither trivial nor an analytical matter as the phase information is squared out in the measured data, and the observables are convoluted by the reference pulse and by the limited spectral resolution of a spectrometer. Given the problem definition (finding an unique polarization pulse form, which, when projected to several angles fulfills measured profiles in the time and frequency domain), the method of choice was a multi-objective evolutionary algorithm.

The first efforts of retrieving the polarization traces by fitting the two goals temporal and spectral projection intensities simultaneously with a Weighted Sum Approach (as used in Chapter 4) caused various difficulties, among them convergence to local instead of global optima, as the combined and convoluted time-frequency search space practically eluded the attempted scalarization to a single goal. Therefore, a

multi-objective algorithm which incorporates the concept of Pareto-optimality was applied for the task. From the various incarnations, a variant implementing sorting of individuals by their Pareto-optimality was chosen, named non-dominated sorting genetic algorithm (NSGA) [161]. It guarantees a continuous fair treatment of the conflicting goals (fitting the temporal and spectral traces) and has practically proved to avoid premature convergence to a single dataset or goal.

In a brief excursion, this chapter first discusses the concept of non-dominant sorting as, to the authors knowledge, it has not yet been incorporated in an OCE; tests an implementation of NSGA-II on a multi-modal problem, then presents the mathematical tools to calculate arbitrary projections before finally combining the methods in an retrieval algorithm

9.2   Multi-objective optimization

Concept of non-domination.   The foundations of multi-objective optimization were set in 1969 in the PhD thesis of Rosenberg [162] who suggested, but not simulated such a procedure. The decisive difference of multi-objective algorithms compared to single objective variants is that not one, scalar fitness per individual is responsible for its advance, but M goals that are pursued simultaneously, which means that the comparison of the individuals for survival has also to be conducted in an M-dimensional "objective space". If an agglomeration of solutions to a single region (called "crowding") is avoided, such an algorithm yields a set of compromise (Pareto-optimal) solutions instead of a single optimal point. This set is said to "dominate" all other individuals, as illustrated in Fig. 4.2. There, four solutions in a 2D objective space were sorted by their usefulness, whereby three of them constituted a "Pareto-optimal" front, and one solution was outperformed by all others, and therefore was "dominated" by them.

The following two conditions must be fulfilled so that individual $x^{(1)}$ dominates $x^{(2)}$ (having $j = 1, \cdots, M$ objectives and more than one objective function f_j, $j = 1, \cdots, M$)[1]

1. $x^{(1)}$ is no worse than $x^{(2)}$ in all objectives: $f_j(x^{(1)}) \not\prec f_j(x^{(2)}) : \forall j = 1, 2, \cdots, M$

2. $x^{(1)}$ is strictly better than $x^{(2)}$ in at least one objective: $\exists j \in \{1, 2, \cdots, M\} : f_j(x^{(1)}) \succ f_j(x^{(2)})$

Sort by domination.   The term non-dominated fronts originates from Chankon [163] (1983), shortly afterwards, a vector-based genetic algorithm (VEGA) was suggested by Schaffer [164] using vectorial fitness values. The first concept of non-dominated sorting was stated in the standard work of Goldberg [42], which was the basis for the two most prominent, elitist multi-objective algorithms used today, NSGA-II [165] and SPEA2 [166]. A non-domination sorting algorithm proceeds iteratively, in a first step determining the non-dominated solutions, assigning their

---

[1]$\not\prec$ and $\succ$ are used so the definition can apply to both maximization and minimization
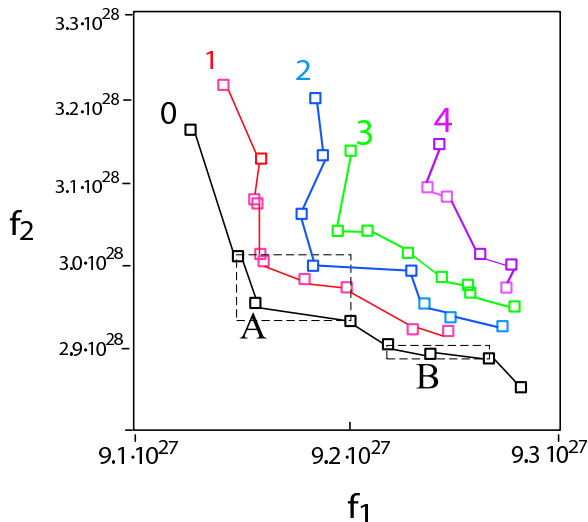
Figure 9.1: A sorting algorithm based on the concept of non-domination determines fronts of equal domination. The dashed rectangle represents the distance of two solutions to their neighbours and serves to assign a crowding distance, which is greater for solution A than for B. In this snapshot, taken from a polarization retrieval, f_1 represents the mean squared error of the fit of all SFG-CC traces and f_2 of the spectra.

domination count, removing them from the set, and sorting the remaining ones again, etc. The assignment can be visualized for two objectives, as illustrated[2] in Fig. 9.1. The lines connect solutions of equal domination count, starting from 0 (non-dominated) (left bottom), working upwards to the right-top corner, as both objectives are to be minimized.

Sort by crowding distance. Early incarnations of multi-objective algorithms used special techniques to avoid a concentration of individuals to single regions (niching) using fitness-sharing [167] of closely spaced individuals. NSGA-II uses a comparably simple, yet less intrusive method using a sorting by crowding "distance". After the solutions were sorted by fronts of equal domination, a second sorting rewards more "lonely" individuals by assigning a crowding distance. This distance is defined as the average side-length of the cuboid determined by the next neighbouring solutions (see the two dashed rectangles in Fig. 9.1). There, solution A is more exposed than solution B, and therefore preferred. For the procedure, the corners are given the highest priority, and then, the rest of the solutions follow by descending crowding distance.

Operators. NSGA-II carries out non-dominated sorting of a pool of parents and offspring, filling every front subsequently with new solutions while keeping the total population size constant. Afterwards, the parents for the next generation are selected deterministically according to their domination rank and crowding distance, before the operators simulated-binary-crossover (SBX crossover) [168] and polynomial mutation [169, 170] are applied.

---

[2]The pulse shapes corresponding to the two corner solutions of the non-dominated front are very similar in structure, meaning that the Pareto-optimal front is indeed very narrow. Still, the emphasis of this particular retrieval is to treat the goals fairly in order to reach an optimal solution.
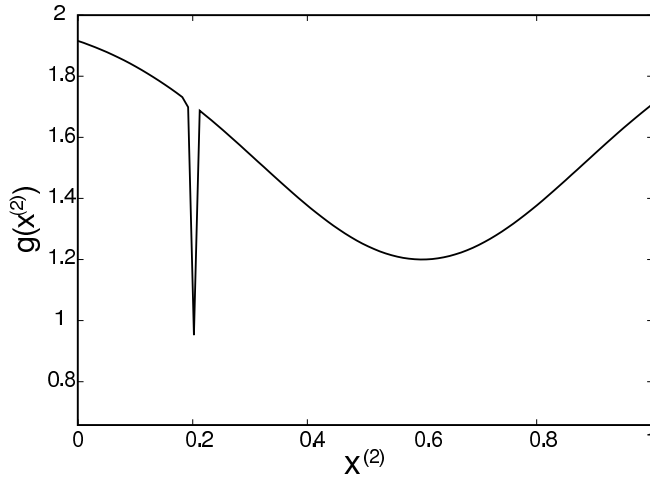
Figure 9.2: Test function $g(x^{(2)})$ with a global minimum at $x^{(2)} = 0.2$ and a local minimum at $x^{(2)} = 0.6$.

9.2.1 Test function for NSGA-II

The retrieval algorithm was first tested on a known function or problem[3], which is a necessary precondition before employing it on real-world problems with unknown solutions. As two objectives were considered for the retrievals (fitting the cross-correlations and the spectra), the chosen test function will also consist of two dimensions, with f_1 and f_2 as the respective fitness components. The easiest test function to think of is $f_1 = x^{(1)}$ and $f_2 = x^{(2)}$ both to be minimized, with a global optimum at $x^{(1)} = x^{(2)} = 0$. More interesting, and also incorporating the commonly encountered trade-off between objectives, a setup like $f_1 = x^{(1)}$ and $f_2 = c/x^{(1)}$ (with only positive values) produces an easily imaginable Pareto-optimal front in objective space, having the form of an hyperbola as $f_1 f_2 = c$. The compromise is easily seen; when the algorithm tries to minimize f_1 by minimizing $x^{(1)}$, f_2 increases due to the denominator $x^{(1)}$. Incorporating a second variable $x^{(2)}$ could, for example, look like

$$
\begin{aligned}
f_1 &= x^{(1)} \\
f_2 &= \frac{g(x^{(2)})}{x^{(1)}}
\end{aligned}
\quad . \tag{9.1}
$$

Constructing $g(x^{(2)})$ in a way that there is a global and a local minimum

$$
g(x^{(2)}) = 2.0 - \exp\left\{ -\left(\frac{x^{(2)} - 0.2}{0.004}\right)^2 \right\} - 0.8\exp\left\{ -\left(\frac{x^{(2)} - 0.6}{0.4}\right)^2 \right\} \tag{9.2}
$$

(taken from from Ref. [171] and plotted in Fig. 9.2) introduces multi-modality and puts an algorithm to a reasonable test. The two Pareto-optimal fronts will have the form (Eq. 9.1), $f_1 f_2 = g(0.2)$ (global), and $f_1 f_2 = g(0.6)$ (local), which also has a larger basin of attraction.

A successful run of the employed NSGA-II algorithm[4] is shown in Fig. 9.3, showing that most of the individuals (green squares) have converged to the global front (black hyperbola), and not to the local (dashed).

---

[3]in the evolutionary computation literature referred to as "experiment"

[4]The objective value's range for $x^{(2)}$ was [0.05;1], an elitist (30,80) algorithm with 15 survivors, σ_s =0.03, using SBX crossover (probability 1, coefficient 12) and polynomial mutation (probability 0.1, no mutational adaptation), deterministic selection; result after 189 generations.
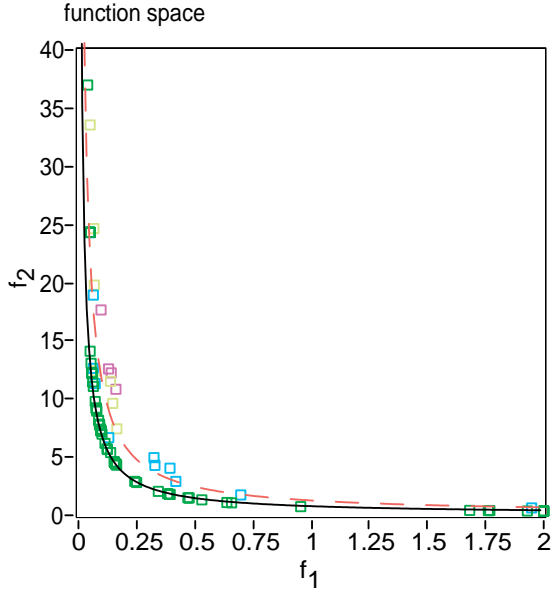
Figure 9.3: Result of a multi-modal test function for NSGA-II. The dashed line is local Pareto-optimal front, the rectangles represent the individual solutions after 189 generations, most of them have already converged to the global front (straight hyperbola), only a few remain elsewhere.

9.3 Pulse retrieval routine

9.3.1 Projected intensities from set B

The procedure of determining the parameters of set E, the intrinsic ellipse parameters, from pulse projections to several angles (used for the TRE-like detection scheme) was described in Sec. 3.4. The retrieval algorithm internally stores the pulse form in the laboratory system, set B: $\{a, b, \varepsilon\}$, so this set's angle α dependent projection (intensity) has to be derived.

By calculating the offset d of the orthogonal ellipse tangent $y = x \tan(\alpha - 90°) + d$ to an axis with the angle α, the projected spectral or temporal intensities $P^2(\alpha)$ must be derived (see Fig. 9.4). The tangent offset d can be calculated solving

$$x^2 \cdot \left(\frac{1}{a \sin \varepsilon}\right)^2 + y^2 \cdot \left(\frac{1}{b \sin \varepsilon}\right)^2 - 2xy \cdot \frac{1}{a\,b} \cdot \frac{\cos \varepsilon}{\sin^2 \varepsilon} = 1. \tag{9.3}$$

(Eq. 7.13) for y, which yields two results (upper and lower ellipse halves),

$$y(x)_\mp = \frac{a\,b\,x\,\cos\varepsilon \mp \sqrt{a^2 b^2 (a-x)(a+x) sin^2\varepsilon}}{a^2} \tag{9.4}$$

where y_- is the lower arc and y_+ the upper arc. Allowing only one solution for x for a given α means solving the equation system of the tangent and Eq. 7.13 for x and y, yielding

$$d = \pm \sin \varepsilon \sqrt{\frac{-b^2 + 2\,a\,b\,\cos\varepsilon \tan\alpha - a^2 \tan^2\alpha}{\cos^2 \varepsilon - 1}}. \tag{9.5}$$

As the projection $P = d \sin \alpha$ is measured as spectral (or temporal) intensity, the last term can be further simplified[5] to

---

[5]The main computational effort for the presented routine is to calculate the field projections (for every data point in time or frequency). For Eq. 9.6, $\cos^2 \alpha$ and $\tan \alpha$ can be calculated once per retrieval run, leaving $\cos \varepsilon$ as the most CPU expensive expression.

Figure 9.4: Calculation of the projection P of an ellipse (determined by the parameters a, b, and ε) to an axis with the angle α.
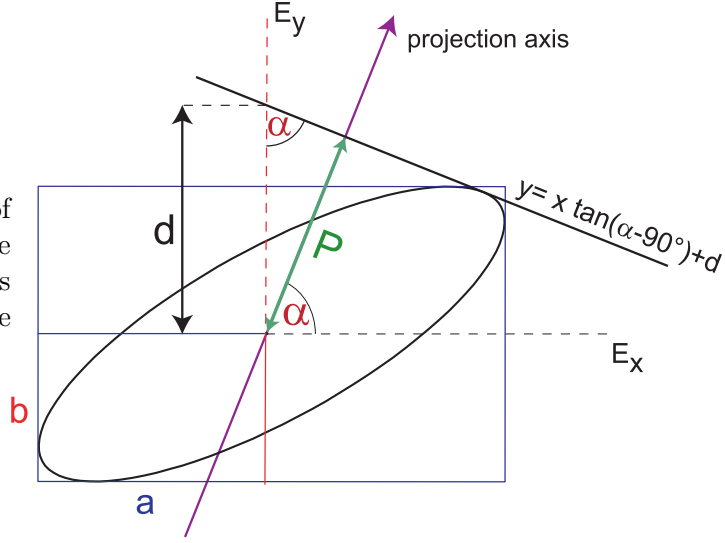
$$P^2(\alpha) = \cos^2 \alpha \left[b^2 + a \tan \alpha \cdot (a \tan \alpha - 2b \cos \varepsilon)\right]. \tag{9.6}$$

To simulate the spectrometer, a convolution (marked by $*$) of the spectral intensity with the resolution function (a spectral Gaussian intensity profile with a width corresponding to the resolution $\delta\lambda$) $I_{res}(\omega)$

$$|E'_{x,y}(\omega)| = \sqrt{|E_{x,y}(\omega)|^2 * I_{res}(\omega)} \tag{9.7}$$

has to be performed, introducing the respective components a' and b'

$$\begin{aligned} |a'(\omega)| &= \sqrt{|a(\omega)|^2 * I_{res}(\omega)} \\ |b'(\omega)| &= \sqrt{|b(\omega)|^2 * I_{res}(\omega)} \end{aligned} \tag{9.8}$$

which transforms Eq. 9.6 into (omitting the absolute value vertical lines)

$$P'^2(\alpha, \omega) = \cos^2 \alpha \left[b'(\omega)^2 + a'(\omega) \tan \alpha \cdot (a'(\omega) \tan \alpha - 2b'(\omega) \cos \varepsilon(\omega))\right]. \tag{9.9}$$

Similarly, the SFG-CCs can be calculated by deriving the cross-correlation with the reference pulse's intensity profile[6] $I_{ref}(t)$

$$|E'_{x,y}(t)| = \sqrt{|E_{x,y}(t)|^2 * I_{ref}(t)}, \tag{9.10}$$

leading to

$$P'^2(\alpha, t) = \cos^2 \alpha \left[b'(t)^2 + a'(t) \tan \alpha \cdot (a'(t) \tan \alpha - 2b'(t) \cos \varepsilon(t))\right], \tag{9.11}$$

meaning that the algorithm automatically includes a deconvolution of the measured temporal and spectral profiles.

9.3.2 Spectral influence of the helicity

The key to retrieving phase and helicity information for a polarization shaped pulse form lies in the (projection) angle dependent spectra $P'^2(\alpha, \omega)$.

---

[6]a temporal Gaussian intensity profile with the respective pulse duration
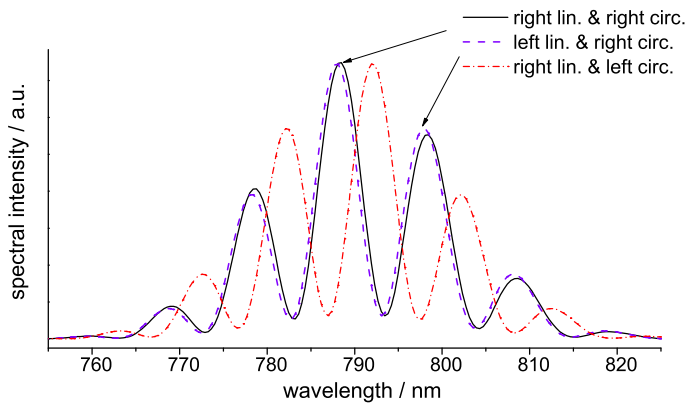
Figure 9.5: Impact of helicity changes on pulse spectra for a 0° linear/circular pulse, calculated for a polarizer angle $\alpha = 67.5°$. Solid lines: helicity swap of the linear pulse, dashed line: circular pulse's helicity swapped from right- to left-handed, which is a comparably large effect.

Fig. 9.5 demonstrates this effect for a $\Delta t = 200$ fs, 0° near-linear/circular double pulse with pulse spectra calculated for a projection angle of $\alpha = 67.5°$. The dashed line results from swapping the circular pulse's helicity, the closely-drawn lines show the difference when the helicity of the near-linear pulse ($H_{ba} = 0.05$) is swapped which, at this angle, shows a smaller effect.

9.3.3 Input data

To demonstrate that the retrieval from SFG-CCs and spectra is functioning as intended, meaning that all the information about polarization and phase is indeed contained, a previously known (simulated) waveform consisting of four sub pulses featuring different intensities, distances, polarization states, and zero order phases (including an overlap of the chirped and the circular pulse) will be restored. Knowing three projections per data point would be sufficient to determine the shape of an ellipse, but to additionally retrieve helicity and phase, more information is required. A set of nine spectra (at least eight, uniformly distributed within an angular interval of 180° are recommended[7]) have shown to be reliable. For the cross-correlations, a minimal number of three is required, having more increases the accuracy of the ellipticity, especially for near-linear cases.

---

[7]to minimize the chance to miss a crucial angle (see Fig. 9.5)

9.3.4 Retrieval algorithm

First, as a reference, the total energy of the waveform has to be determined, integrating

$$\int P_k'^2(\alpha, \omega) d\omega = A_k^2(\alpha) \tag{9.12}$$

over all frequencies or alternatively, integrating $P_k'^2(\alpha, t)$ over all times, for all available k spectral or temporal traces. Then, a sinusoidal function is fitted through $A_k^2(\alpha)$ as

$$A'(\alpha) = a_0 + a_1 \cdot \sin(2\alpha + a_2) \tag{9.13}$$

(similar to Eq. 7.39, or as shown in Fig. 7.7) in order to yield a continuous function of α, and thus, the energy contained in the waveform is $2a_0$. The optimal solution, a filter $\tilde{\mathbf{H}}_{opt}^{par}$, must therefore provide the same $P'^2(\alpha, \omega)$ and $P'^2(\alpha, t)$ traces as the input data, if multiplied with a reference field $E_{in}(\omega)$ of the energy $2a_0$.

Fig. 9.6 shows a sketch of the evolutionary retrieval algorithm and explains the retrieval procedure in the figure caption. It turned out that the NSGA-II operators simulated-binary-crossover (SBX crossover) [168] and polynomial mutation [169, 170] could be exchanged for their computationally less expensive, standard versions. The mutation rate was adapted with a 2/5 decision factor to the best progress of either time- or frequency domain, depending on how good the respective offspring performed in either domains. Furthermore, a non-elitist (12,40)-strategy could be used instead of the elitism of NSGA-II. The search space sampling was 512 points, the spectral sampling resolution was 0.3 nm, the temporal resolution 14 fs, $I_{res}(\omega)$ was 0.55 nm, and the reference pulse's FWHM was 25 nm.
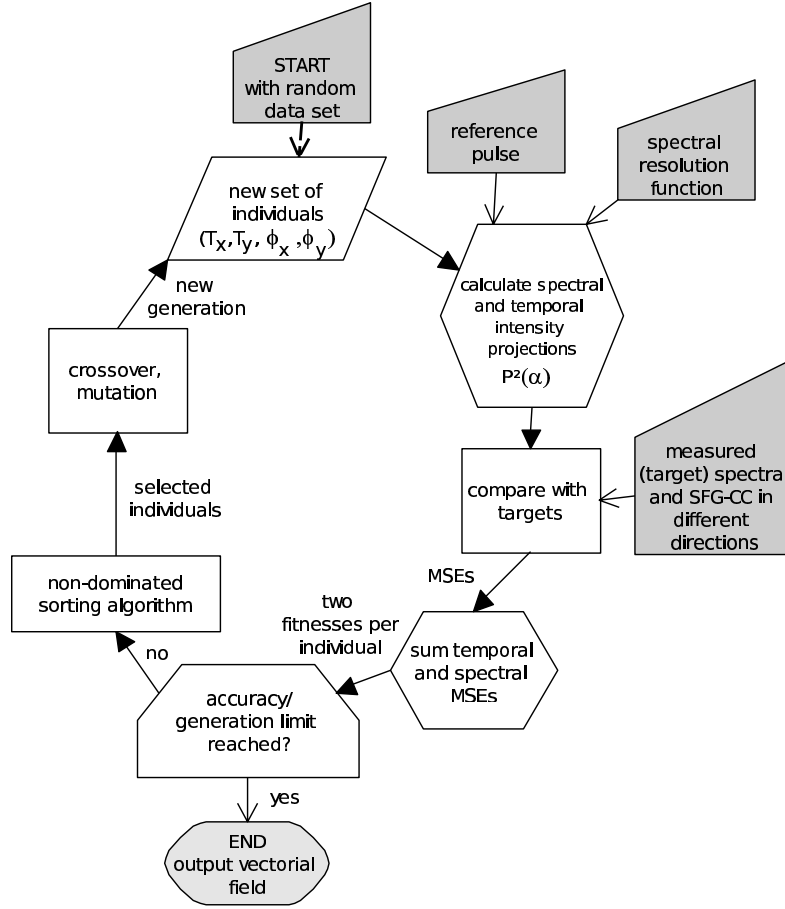
Figure 9.6: Flowchart diagram of the retrieval algorithm. The encoding is set in the frequency domain, a chromosome consists of the four data strings T_x, T_y, ϕ_x, and ϕ_y; which also comprise the general filter function $\tilde{\mathbf{H}}(\omega)$ also used to describe the parallel setup in Chapter 8. The algorithm starts with a random initial guess in order to obtain the first generation's set of 40 individuals. In the next step, using the information encoded in the genes, the transfer function $\tilde{\mathbf{H}}(\omega) = \hat{x}\sqrt{T_x(\omega)}e^{i\phi_x(\omega)} + \hat{y}\sqrt{T_y(\omega)}e^{i\phi_y(\omega)}$ (see also Eq. 8.1) is constructed. Again, the output field is determined as $\mathbf{E}'_{out}(\omega) = \tilde{\mathbf{H}}(\omega) \cdot E_{in}(\omega)$. Afterwards, the spectral projection intensities $P'^2(\alpha, \omega)$ of $\mathbf{E}'_{out}(\omega)$ (Eq. 9.9) are calculated using Eq. 9.8, where the x- and y-components are denoted as a and b, respectively. For the simulated SFG-CCs, the spectral field $\mathbf{E}'_{out}(\omega)$ first has to be inversely Fourier transformed to become $\mathbf{E}_{out}(t)$, before calculating $P'^2(\alpha, t)$ using Eqs. 9.10 and 9.11. In order to obtain the two fitnesses f_1 and f_2, the mean squared errors of $P'^2(\alpha, \omega)$ and $P'^2(\alpha, t)$ are added to their corresponding measured traces as $f_1 = \sum_\alpha MSE_{\alpha, temp}$ and $f_2 = \sum_\alpha MSE_{\alpha, spec}$, resulting in two overall fitness "dimensions" which were both to be minimized. The non-dominated sorting algorithm performs the ranking according to the degree of non-domination and crowding distance. Afterwards, the 12 most fit individuals (from the 40 original) are deterministically selected (in the order how they were sorted), undergo crossover in order to produce 40 new offspring for the next iteration, which then are mutated. If a sufficient coincidence or a fixed number of generations is reached, the algorithm finishes and yields the output filter $\tilde{\mathbf{H}}^{sol}(\omega)$.
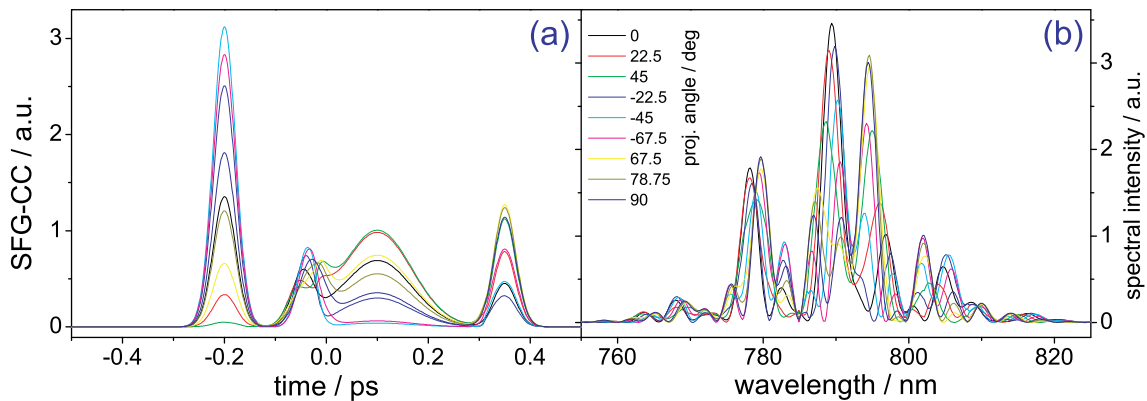
Figure 9.7: SFG-CC (a) and spectral intensity (b) projections to various angles serve as input for the algorithm, showing a waveform consisting of four, parametrically generated sub pulses.

9.3.5 Quadruple pulse retrieval

To demonstrate that the method is functioning as intended and there is enough information contained in the projections, a previously known waveform is retrieved, which is a parametrically constructed quadruple pulse. It consists of four sub pulses which features different intensities, distances, polarization states, and zero order phases; including an overlap of the chirped and the circular pulse.

Fig. 9.7 states the input for the algorithm in the time (a) and frequency (b) domain. The respective traces serve as reference in order to calculate the mean squared error of the current individual's temporal and spectral projections, whereby the projection spectral intensities (b) show to be highly structured.

Fig. 9.8 shows the result of the retrieval, which was obtained after 5000 iterations[8]. The straight line represents the original parameter courses over time, and the dashed lines the retrieved values. Graph (a) shows the temporal intensity, (b) the ellipticity H_{ba}, (c) the major axis angle $\beta/2$, (d) the helicity, and (e) the sum-phase $\phi_{sum} = \phi_x + \phi_y$, where the linear chirp of one of the sub pulses can be qualitatively spotted as a quadratic shape. A very good agreement could overall be obtained even with the reduced sampling; it is also apparent that the coincidence scales with intensity. The helicity could be well recovered for all states which were not too close to linear polarization. The result for the sum-phase (e) is not unambiguous, it bears a constant phase offset which is of no practical consequence for non phase-stable laser systems, and was manually adapted for the comparison.

Retrieval course. It is interesting[9] to observe the multi-objective algorithm's behavior during the run and to point out features which enables it to treat the two goals fairly.

Fig. 9.9 shows the changing internal states over generation for the quadruple pulse retrieval. In (a), the progress for the two individual goals f_1 (temporal fit/ black) and f_2 (spectral fit - red/ gray) is shown. The risk that scalarization methods face, namely favoring one goal over the other, does not apply due to non-dominated

---

[8]about 40 minutes of computing time on an AMD Athlon XP 2200+

[9]in anticipation of upcoming experiments
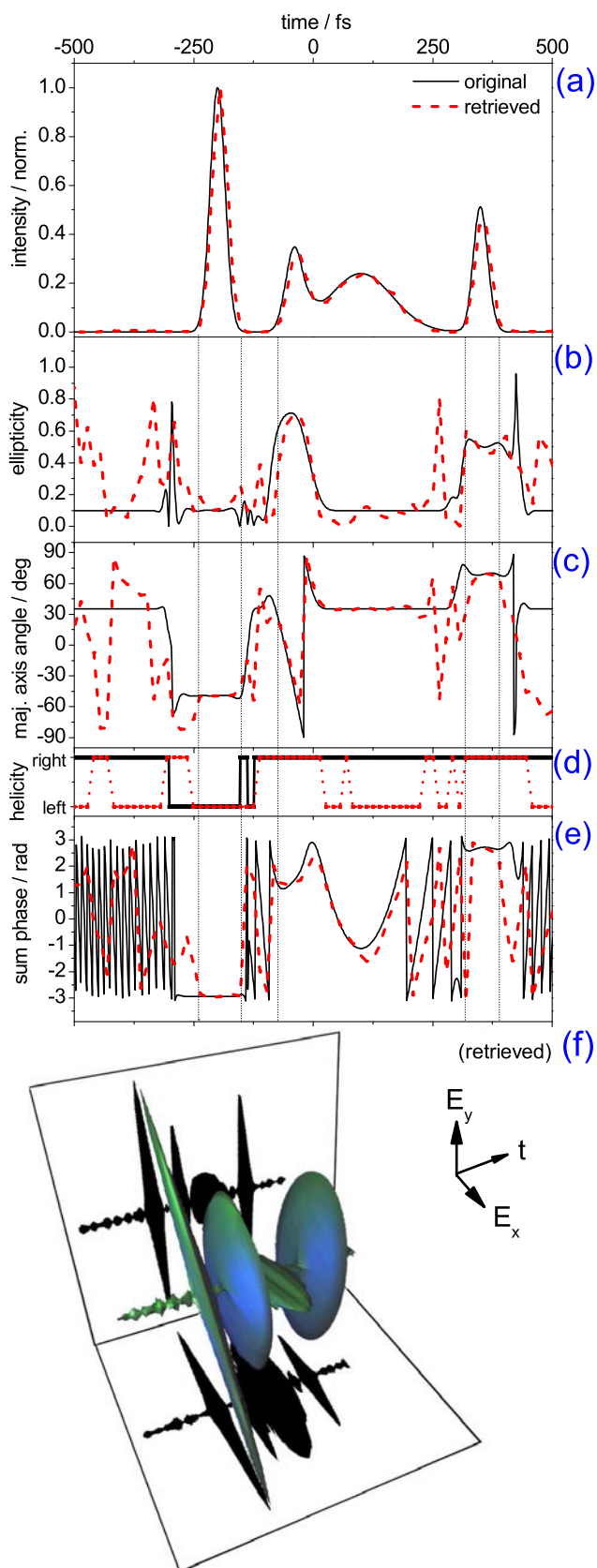
Figure 9.8: Target and retrieved pulse form, using only spectra and SFG-CCs.
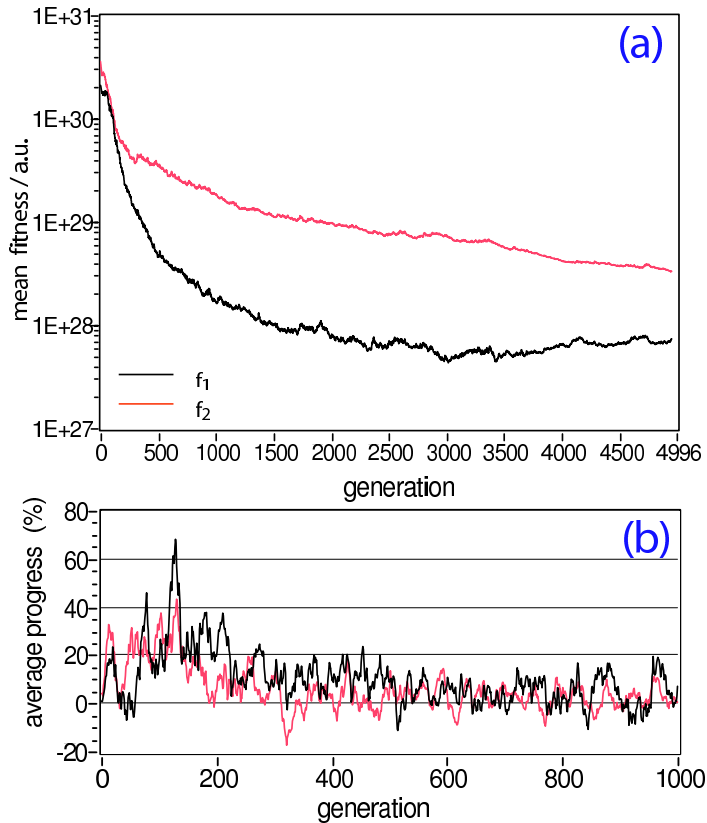
Figure 9.9: Inner workings of the retrieval algorithm. (a) Evolution of the two respective goals f_1 (SFG-CCs) and f_2 (spectra); (b) shows the relative progress in fitness relative to 20 generations before.

sorting, which can also be seen by the fact that the SFG-CC fits f_1 improve considerably at the beginning, while having lower (arbitrary) values than f_2. In Graph (b), the initial, relative progress of the two goals relative to the values 20 generations before (averaged over three generations) is depicted. The graphs illustrate an important feature of a multi-objective algorithm, a "sacrifice" of one progress rate towards another; whereby the progress of one goal comes to halt or even deteriorates, in order to allow the other to "flourish".

9.3.6 Experimental pulse retrieval

Finally, an experimental demonstration incorporating, unfortunately, only four SFG-CCs and spectra[10] is presented. It shows a successful convergence to all four spectral and four temporal goals under experimental, noisy conditions, without claiming a successful retrieval of the helicity and phase.

The experimental pulse retrievals (see Fig. 9.10) showed no particular degradation of fit quality by typical measurement noise, as an evolutionary algorithm innately can overcome noise to a certain degree. The individual fit results after 13000 generations showed a good convergence for all input angles (-20°, 0°, 45°, and 90°) when comparing the SFG-CCs (left column) and spectra (right column).

---

[10]as it was still not yet clear at the time of data acquisition that more that four spectra would be necessary for a reliable retrieval
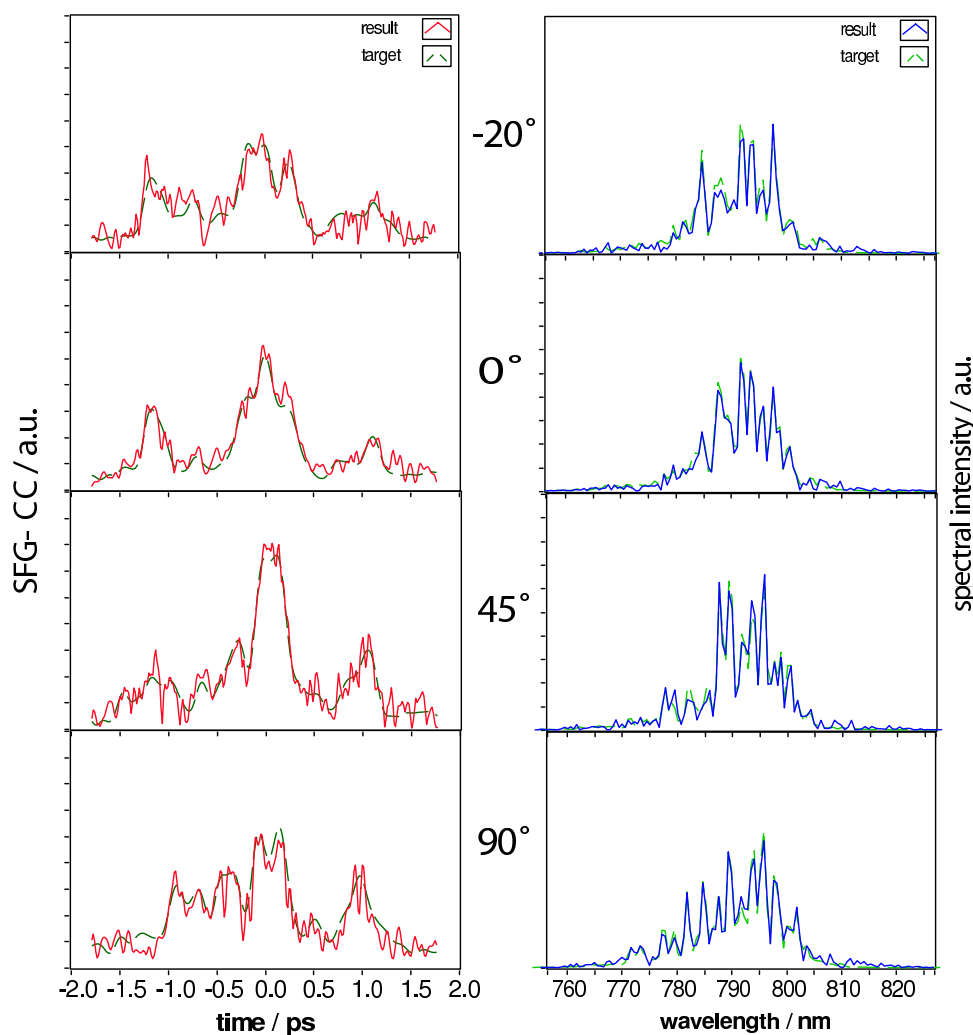
Figure 9.10: Retrieval of an experimentally generated pulse form. Left column: SFG-CCs, right column: spectral intensities; green/dashed lines: targets, straight lines: retrieval results. Top to bottom: -20°, 0°, 45°, and 90° projections.

9.4   Summary and Outlook

In this chapter, a method capable of fully retrieving polarization shaped pulse forms which does not require an interferometric detection, was presented. A retrieval for a complex pulse shape demonstrated that there was enough information embedded in a set of cross-correlations and spectra for this task. From an experimental point of view, it is not too costly (a set of SFG-CC traces and spectra - behind a polarizer - have to be measured), and the computational effort employing a multi-objective evolutionary algorithm is also manageable, as such algorithms are widely (and freely) available [172, 173, 174].

As for an outlook, multi-objective algorithms, to the author's knowledge, are not used in femtochemistry experiments so far[11], only scalarization methods seem to be employed on a regular basis [55, 85, 86, 87, 88, 89] but carry with them the problems which non-Pareto approaches bring along, which were discussed throughout

---

[11]only a draft of a theoretical study [91], and analysis of RAMAN-spectra [175] could be found

this work. Using Pareto-methods and employing multi-objective algorithms instead would allow to perform experiments with the advantage not only not to be bothered with objective weightings, but also obtaining a set of solutions that can shed light onto the causes that underlie processes and allowing to investigate correlations between conceptually arbitrary aims. That way, a systematic assignment how specific features favor or facilitate certain goals could be conducted. For example, the relation between ionization versus ionization plus fragmentation of molecular systems could be studied in order to find a Pareto-optimal front which contains pulse forms that could comprise features (sub pulses, polarization states, peaks in spectra, etc.) which could be traced along the front in order to find a decisive feature change leading to the different molecular dynamics. For example, the issue of fragmentation of more complex molecules into various daughter ions could be explored in a systematic way - even with a higher dimensionality - by optimizing three or more fragments (with perhaps even different isotopic masses) simultaneously. Furthermore, the shape of the Pareto-optimal front itself could be used to study (non)linearities of the involved processes.