LASER PULSE DURATION OPTIMIZATION WITH NUMERICAL METHODS

Francesco Capuano[1,*], Davorin Peceli, Bedřic Rus, Alexandr Špaček

ELI, Institute of Physics, Czech Academy of Sciences, Dolní Brežany, Czech Republic

Gabriele Tiboni[†], Politecnico di Torino, Turin, Italy

[1]also Politecnico di Torino, Turin, Italy

Abstract

In this study we explore the optimization of laser pulse duration to obtain the shortest possible pulse.

We do this by employing a feedback loop between a pulse shaper and pulse duration measurements. We apply to this problem several iterative algorithms including gradient descent, Bayesian optimization and genetic algorithms, using a simulation of the actual laser represented via a semi-physical model of the laser based on the process of linear and nonlinear phase accumulation.

INTRODUCTION & RELATED WORKS

L1-Allegra System

L1 Allegra system is high power laser system developed in ELI Beamlines in Czech Republic designed to deliver <20fs pulses with energy higher than 100 mJ at a repetition rate of 1 kHz. This system is based on the amplification of frequency-chirped pico-second pulses in an Optical Parametric Chirped Pulse Amplification (OPCPA) chain consisting of seven amplifiers. L1 contains 7 OPCPA stages that are pumped by 5 diode pumped lasers based on commercial Ybdoped thin-disk Regenerative Amplifiers (RA) DIRA 200-1. All of the pump lasers generate pulses at 1030 nm with 1 kHz repetition rate and also include stretcher, compressor and an LBO crystals for second harmonic generation (SHG) of pulses at 515 nm. To achieve high SHG efficiency several features of the pump laser system should be optimized. One of these features is the laser pulse temporal shape. The optimization of temporal pulse shape is accomplished through the manipulation of the pulse spectral phase. A standard method for spectral phase representation employs a Taylor expansion of spectral phase, $\varphi(\omega)$, around the central angular frequency ω_0, of the spectral pulse:

$$\varphi(\omega) = \sum_{=0}^{\infty} \frac{\partial^{(\,)}\phi}{\partial\omega^{(\,)}} \cdot \left(\omega - \omega_0\right) .$$ (1)

The first two terms of the sum of Eq. (1) represent constant phase and group delay and don't have effect on the pulse temporal shape. Third, fourth and fifth term are instead related to the dispersion parameters group delay dispersion (GDD $\sim \varphi'$), third order dispersion (TOD $\sim \varphi''$) and fourth order dispersion (FOD $\sim \varphi'''$) and do have an impact on the temporal profile. In this work, we aim at control these parameters, to which we will collectively refer to as ψ.

---

* Francesco.Capuano@studenti.polito.it
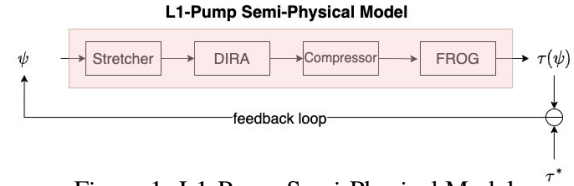† Gabriele.Tiboni@polito.it

Figure 1: L1-Pump Semi-Physical Model.

Since SHG is an intensity-dependent non-linear optical process, the highest conversion efficiency is obtained for those pulses with temporal shape closest to transform-limited pulses. In particular, since the transform-limited (TL) pulse is the pulse with minimum possible pulse duration given the present spectral bandwidth, it is guaranteed to have the highest possible pulse intensity.

Temporal properties of amplified pulses are controlled through manipulation of parameters of the stretcher by *Teraxion* and monitored using SH FROG 1030 by *Femtoeasy*. The control of pulse temporal shape is done by manipulating three parameters, d_2, d_3 and d_4 on the stretcher controls that relate to parameters GDD, TOD and FOD previously introduced through a determinate system of linear equations, depending on the central wavelength (derived from central frequency) of the spectrum considered.

Related Works

Obtaining short and sharp laser pulses is crucial for running high-efficiency physical systems that would reach really high intensities with limited energy consumption.

Various algorithms were used in ultra-fast laser systems for active feedback optimization of laser parameters or laser-matter interactions [1, 2], but none of these previous approaches optimized the single pulse shape.

In this work we compare three different algorithms for the optimization of the temporal profile of laser pulses. These are: Bayesian Optimization, Differential Evolution, and a custom implementation of gradient descent.

Semi-Physical Model of L1-Pump-System

Optimization of laser pulse shape is designed through feedback control loop between stretcher and FROG. To test performances of different optimization algorithms we designed a theoretical semi-physical model of the L1 amplifier chain containing stretcher, DIRA, compressor and FROG, as presented in Fig. 1.

Each of the elements in the pump chain contribute to the overall spectral phase. Furthermore, while stretcher and compressor introduce only linear change to the spectral phase (GDD, TOD and FOD), DIRA also introduces nonlinear

effects. The system presented in Fig. 1 is parameterized by: **B-integral** (representing a measure of the accumulation of light's nonlinear phase), estimated to be 2 for our DIRA; **Compressor parameters**, considered constant in this work and estimated as opposite to the nominal values of stretcher's parameters.

We conclude the presentation of our semi-physical model highlighting how DIRA's contribution in terms of linear spectral phase can be neglected.

PULSE OPTIMIZATION

Other works did tackle the problem of laser optimization focusing on using Deep Reinforcement Learning to obtain mode-locking inside the laser cavity [2]. However, no previous work was done to optimize the actual pulse shape of single pulses so as to make it as similar as possible to an arbitrary target shape.

In this work (freely available on GitHub [3]), we focus exactly on this problem. If one indicates with τ^* such target shape and with $\tau(\psi)$ the temporal shape resulting from the adoption of the control ψ on the system, then it is possible to frame the problem as:

$$\min_{\psi \in X} L(\tau^*, \tau(\psi)) \, , \qquad (2)$$

where $L : \mathbb{R}^n \mapsto \mathbb{R}^+$ is a figure of merit which is used to guide the optimization route towards better and better values of ψ and X is the region of the parameters space containing physically-feasible configurations.

In this work we adopted as target temporal profile the transform-limited pulse, i.e. the temporal representation of the pulse corresponding to a zero phase imposed on the original spectrum.

In the case in which the system presented in Fig. 1 were not present any non-linear blocks such as regenerative amplifiers, this result can be obtained imposing a control which corresponds to $\psi_{str.} = -\psi_{compr.}$. However, since this nonlinearity is indeed present in the system (as the beam energy needs to be increased), it is not possible to obtain the transform-limited pulse applying such control.

This can also be seen in Fig. 2, where it is possible to see that the non linearity present heavily distorts the output of the application of an otherwise optimal control.

Loss Function

We shall first observe that an exact analytical solution to the problem of mixed (linear and non-linear) phase accumulation it is practically impossible to obtain in our case.

This fact generates the need of adopting a solution that, with an iterative approach, adjusts the control parameters so as to find a region in which the control induces a temporal profile matching the desired one under an acceptable threshold. Such a threshold is computed with respect to an arbitrary, yet very relevant, L.

Our first guess has been the Mean Squared Error (MSE) loss function. However, we observed that this loss function is

really not informative as a guiding signal for the optimization process. This is due to two main reasons:

- Since the temporal profile is represented via a 0-1 normalization, differences are also in this range. This implies that differences in the order of magnitude of one tenth -which are very significant in this case- account only for values in the order of magnitude of one cent, when squared (according to MSE formulation). Since the majority of the two curves (the tails of the impulse) numerically assumes very similar values, the MSE's numerator is very likely to be a small number.

- The algorithm we used to perform Fourier Transformations requires a very large number of points (often 30k) in order to obtain temporal profiles with acceptable resolutions. This essentially means the MSE's denomiator would always be a large number, independently on the actual temporal shapes, thus biasing the final a value towards small values.

In light of this, we used different loss functions for our problem, as MSE would be inevitably biased towards small numbers and, therefore, not a valuable guiding signal in any optimization route.

Optimization Algorithms

To obtain the configuration of parameters which minimizes one of the losses considered, we implemented various optimization algorithms. In particular, we implemented three algorithms: two so-called "black box" algorithms, namely Bayesian Optimization and Differential Evolution, and one "first order" algorithm based on first-order differential information employed in an implementation of the ADAGRAD algorithm [4].

It is important to note that while both the black-box algorithms can, in principle, be applied to the real-world laser in the sake of optimizing its configuration, ADAGRAD cannot. ADAGRAD could be applied approximating if the gradients were to be approximated with finite differences, but this would very much increase the computational effort required.
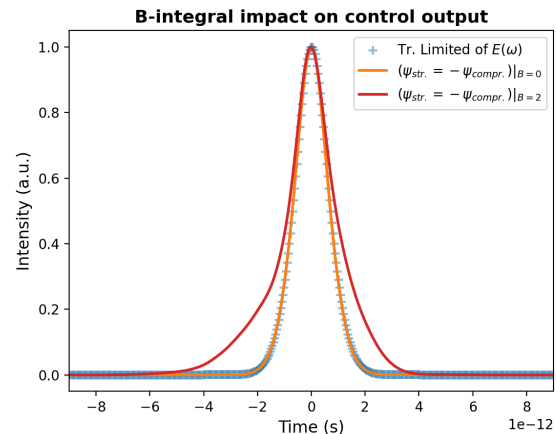

Figure 2: Impact of B-integral and control parameters on output shape.

However, since we plan on collecting enough real data to train an Articial Neural Network (NN) to reconstruct the $\psi \mapsto \tau$ mapping (which would support automatic differentiation), we decided to experiment with this algorithm despite its immediate lack of practical applicability.

Bayesian Optimization   Bayesian Optimization is a popular black-box optimization algorithm that probes at random various points inside a feasible region $\mathcal{X}$, adjusting, according to Bayesian Theorem, a probability distribution on the function fitting best those points.

This function is referred to as "surrogate function" or, more commonly, "acquisition function". At each iteration, the new point to be probed is selected balancing between exploration of the search space and maximization of this acquisition function.

We want to stress that this algorithm should not be directly applied on the real hardware, as the exploration of the search space causes serious oscillations between subsequent probed points, thus posing the actual dynamical system under a large amount of stress.

As a final note, it is important to note that in our implementation we restricted the number of iterations to 150, as we observed performance improvement stagnation over around 120 iterations and obtained a result presented in Fig. 3. As it is possible to see, the obtained results presents some wings that make it significantly different from the target profile for what concerns pulse mounting and de-mounting. Those wings could probably be by further exploring the GDD and TOD search space.

Among all the different loss functions tested, the one that gave the best results is the **sum-L1-Manhattan norm**. We used the default acquisition function-exploration coefficient (κ) configuration of [5] as hyper-parameters.

Differential Evolution   Like Bayesian Optimization, Differential Evolution (DE) is a black-box algorithm that explores a unknown landscape in a feasible region $\mathcal{X}$.

This algorithm mimics the process of natural selection, initializing a population of candidate solutions that iteratevely gives birth to mixed individuals.

The fitness of the various candidate solutions is typically strictly related to the value of the objective function evaluated at each individual point. The algorithm is defined so as to promote some characteristics of the best individual in each generation's offspring, so as to have that as generation go by, the fitness level increases as well as the quality of the candidate points.

While with this algorithm we observed a significant decrease in oscillations between subsequent points (particularly at the beginning), we must observe that such oscillations are most likely not very dangerous for the actual system, since they are very rare.

However, this solution may be really slow when applied to the actual machinery. Such a increase in the needed time is mainly due to the fact that, when in the real world, each function evaluation lasts at least as the actual physical process underlying the control application. Since DE is a method which requires a high number of function evaluations, it is clear how this method could turn out to be really slow as minutes may be necessary to complete only one of a multitude of necessary iterations, especially if starting in a region that eventually turns out to be far apart from the optimal point.

Among all the different loss functions tested, the one that gave the best results is a mixture of one which computes the **MSE only for non-zero values**, (thus discarding the pulses' long tails) and a measure of **the mismatch in terms of subtended area by controlled and target pulse** (weighted in favour of difference of area, respectively with 0.3-0.7 weights).

As a final note, it is important to note that in our implementation we restricted the number of iterations to 150 and the population size to 20. The mutation coefficient, responsible to combine population's individuals has been set to 0.8 whereas the recombination probability between the best candidate and the mutant has been set to 0.7. The results obtained with this algorithm turned out to be far better than the ones obtained with Bayesian Optimization and are presented in Fig. 3. It is possible to see a pedestal which indicates a sub-optimal exploration of the TOD and FOD space, but the obtained pulse is very much similar to the transform limited for the considered set of parameters.

ADAGRAD   Unlike the two other proposed approaches, ADAGRAD assumes the possibility of accessing first-order differential information so as to build an optimization algorithm which ultimately differs from regular steepest descent for its capability of adapting and scaling its step-length with respect to the dimension it is optimizing.

This is particularly interesting for our setting, as the feasible region $\mathcal{X}$ presents some peculiarities, such as the fact that between the first and the last dimension there are almost than 30 orders of magnitude. Therefore, an adaptive strategy for the learning rate must be put in place.

ADAGRAD was born as an unconstrained optimization algorithm, and has had a wide adoption in NN-training. To decline this algorithm to this constrained case we employed the-
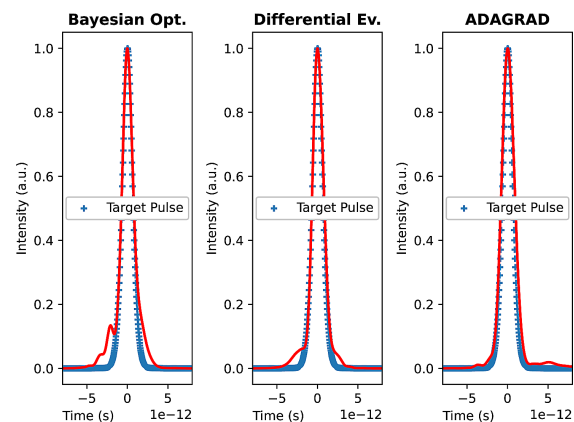

Figure 3: Final results obtained with different algorithms.

Table 1: Final Results Obtained

Algorithm	Number of func. evaluations	Final MSE reached
Bayesian Opt.	150	2.625×10^{-5}
Differential Ev.	200 (its) $\times$ 20 (pop-size) = 4000	1.450×10^{-5}
ADAGRAD	3000 (its) = 3000	4.899×10^{-6}

oretical results presented in [6], such as Sequential Penalty Methods, turning the constrained optimization problem into a constrained one.

At each iteration, the optimization route is directed in the steepest descent direction using an adaptive learning rate which re-scales an initial learning rate η_0 using the inner product of all past gradients so far. This helps scaling the step-length using information on past steepest directions which, ultimately, leads to an adaptive learning rate. If the new point is outside of the feasible region, then the objective function is penalized with an increasing penalty terms that eventually pushes the optimization route into $\mathcal{X}$.

This algorithm is definitely the safest in terms of relative dissimilarity between two different candidate points (since this difference can be controlled using η_0) and the one requiring the smallest number of function evaluations. However, it is practically deployable only in the case in which a NN able to approximate the mapping $\psi \mapsto \tau$ is built. Alternatively, this method could be built using finite differences, thus resulting in eliminating the need for the ANN. In this case, the number of function evaluations could be contained to be even one order of magnitude smaller than the number of function evaluations needed in DE, for instance.

Among all the different loss functions tested, the one that gave the best results is the **natural logarithm of the sum-L1-Manhattan norm**. We used an exponentially increasing penalty term equal to $\frac{1}{10\cdot(0.9999)^k}$ and an initial learning rate of $\eta_0 = 10$. Moreover, we mapped ψ to its $\mathrm{fs}^2, \mathrm{fs}^3$ and fs^4 representation to improve the numerical stability of the algorithm.

The results obtained with this algorithm are presented in Fig. 3. As it is possible to see, ADAGRAD practically reproduces the transform-limited pulse, although the right shoulder of the pulse is a bit off. This is most definitely due to an exploration of the GDD dimension, which we plan on further analysing soon.

Results

The algorithms just presented have been all used to optimize the phase so as to obtain the most similar shape to transform-limited. It is worth noting that, while we did

benchmark our model and observed that it captures some fundamental trend in actual FROG-reconstructed pulse shapes, we did limit the complexity of the model to keep its run-time in bearable time for algorithms employing a typically large number of iterations.

Our results are presented in terms of obtained shapes in Fig. 3 and, for what concerns the final MSE reached, in Table 1.

CONCLUSION

In this work we showed how to effectively use different algorithms to modify the temporal shape of the L1-Allegra high-power laser pulse. The modification of the temporal profile has a wide range of applications, as it is a first step towards to reach the high intensities that users and researchers at ELI Beamlines demand. To obtain such result, we did explore the space of parameters with three different supervised algorithms, as they all needed a target temporal profile to carry out the parameters optimization.

However, we did observe a high dependency of the results found (respectively, the best configuration of parameters obtained with the different algorithms) on the parametrization of laser environment. This is a crucial problem, as precise knowledge about the laser's parametrization is typically non-obtainable.

As future work, we plan on expanding the Pulse Optimization process with techniques which are robust to such a problem. In particular, sequential decision making approaches -such as in Reinforcement Learning- may be investigated to achieve adaptive, online control of high-power lasers that further complies with real hardware constraints on machine safety.

REFERENCES

[1] T. Baumert, T. Brixner, V. Seyfried, M. Strehle, and G. Gerber, "Femtosecond pulse shaping by an evolutionary algorithm with feedback," *Appl. Phys. B: Lasers Opt.*, vol. 65, no. 6, 1997. doi:10.1007/s003400050346

[2] E. Kuprikov, A. Kokhanovskiy, K. Serebrennikov, and S. Turitsyn, "Deep reinforcement learning for self-tuning laser source of dissipative solitons," *Sci. Rep.*, vol. 12, no. 1, pp. 1–9, 2022. doi:10.1038/s41598-022-11274-w

[3] *Eliopt.* https://www.github.com/fracapuano/ELIopt

[4] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, no. 7, pp. 2121–2159, 2011. http://jmlr.org/papers/v12/duchi11a.html

[5] F. Nogueira, "Bayesian Optimization: Open source constrained global optimization tool for Python," 2014. https://github.com/fmfn/BayesianOptimization

[6] S. Wright, J. Nocedal, *et al.*, "Numerical optimization," *Springer Science*, vol. 35, no. 67-68, p. 7, 1999.