Active Muon Shield for the SHiP Experiment at CERN

A. Baranov[1,2,⋆], E. Burnaev[3], D. Derkach[1,⋆], A. Filatov[1,⋆], N. Klyuchnikov[3], F. Ratnikov[1,2,⋆], A. Ustyuzhanin[1,2,⋆] and A. Zaitsev[3]

[1] National Research University Higher School of Economics, Moscow, Russia
[2] Yandex School of Data Analysis, Moscow, Russia
[3] Skolkovo Institute of Science and Technology, Moscow, Russia
[⋆] on behalf of the LHCb Collaboration

E-mail: filatovartm@gmail.com

Abstract. The SHiP experiment is designed to search for very weakly interacting particles beyond the Standard Model which are produced in a 400 GeV/c proton beam dump at the CERN SPS. An essential task for the experiment is to keep the Standard Model background level negligible. In the beam dump, around 10^{11} muons will be produced per second. The muon rate in the spectrometer has to be reduced by at least four orders of magnitude to avoid muon-induced backgrounds. It is proved that novel active muon shield may be used to magnetically deflect the muons out of the acceptance of the spectrometer.

1. Introduction

The Standard Model of elementary particles despite its enormous success has several places for improvement [1]. In particular, the baryon asymmetry of the Universe [2] and a Dark Matter existence are yet to be explained. Heavy Neutral Leptons (HNLs) [3], which are right-handed partners of the Standard Model neutrinos should they be discovered will give a boost in our understanding of both above mentioned riddles of Nature. The existence of such particles is strongly motivated by theory and were searched extensively previously. Cosmological constraints on the properties of HNLs now indicate that the majority of the interesting parameter space for such particles was beyond the reach of the previous searches at the PS191 [5], BEBC [7], CHARM [4], CCFR and NuTeV [6] experiments.

For this reasons, a new fixed-target experiment at the CERN SPS accelerator was proposed [8]. This experiment will use a 400 GeV proton beam on a fixed target to produce a large number of charm mesons. The HNLs from charm meson decays have a significant polar angle with respect to the beam direction, approximately 50 mrad on average. In order to maximise the geometric acceptance for a given transverse size of the detector, the detection volume must therefore be placed as close as possible to the target. The production of the charm mesons is accompanied by copious direct production of pions, kaons and short-lived light resonances. The subsequent decays of these particles would result in a large flux of muons and neutrinos. To minimise these decays, a combination of a target and a hadron absorber of a few metres length, both made of as dense a material as possible, is required. To reduce the detector occupancy and backgrounds induced by the residual muon flux, a muon shield is required downstream of the

hadron absorber. The experimental set-up must therefore balance the opposing requirements of locating the detector as close as possible to the target and of accommodating a sufficiently long muon shield upstream of the fiducial volume of the detector to reduce muon-induced backgrounds.

2. Bayesian Optimization

The main goal of our research was to find a light and efficient shield. In order to achieve this, we applied a Bayesian optimization algorithm. In this section we give an introduction to this method and motivation behind that approach.

Let $f(x)$ be a black-box function on $X \subset R^d$, which does not have an analytic expression and can be evaluated at a specific point only by conducting computationally intensive simulations, which in turn can yield noisy observations of the form $y_i = f(x_i) + \epsilon_i$, where ϵ_i is the random variable with zero mean and constant variance. Our ultimate goal is to find a point x_* where $f(x_*)$ is a true global optimum (henceforth without loss of generality we will consider a maximisation problem). In this problem setup, the general framework for optimum searching iteratively constructs the sample $D_N = \{(x_i, y_i)\}_{i=1}^N$, each iteration consists of three steps:

(i) Choose a new point x_i for simulation

(ii) Conduct simulations to obtain $y_i = f(x_i) + \epsilon_i$

(iii) If a stopping criterion is not met, add (x_i, y_i) to the sample D_N and proceed to iteration $i + 1$

Convergence to the exact solution of the problem may require infinite amount of iterations, however, under some theoretical assumptions any degree of its approximation can be obtained within the corresponding finite horizon.

Theorem 1

If $f(x)$ is Lipschitz-continuous with a Lipschitz continuity constant C, X is a d-dimensional unit hypercube, then in the noise-free case for any $\varepsilon > 0$ some point x_+ with near-optimal true value $f(x_+) : f(x_+) \geq \max_{x \in X} f(x) - \varepsilon$ is guaranteed to be found within $(C/2\varepsilon)^d$ iterations. (see [10])

Strong convergence guarantee is provided against the worst-case scenario, however the solution in this case will typically require an impractically large number of simulations for any reasonable accuracy of approximation. Since such scenarios are not plausible in practice and the number of simulations is very limited, next idea that comes is to relax guarantees in order to meet practical limitations, yet insure against average-cases.

Bayesian approach to optimization allows implementing the idea of guarantees relaxation for the sake of practical applicability by providing specific choosing strategies (step i) in the framework for optimum search. In this approach, the objective function is approximated with *a surrogate model*, that is, a parametric function for which prior knowledge of parameters combined with likelihood of the observed sample defines a posterior distribution over the parametric family. Choice of new points to simulate is guided by *the principle of minimum expected risk* (its formalisation under the Bayesian framework has been provided in [9]) resulting in a secondary optimization problem, which in turn depends not on the original function, but on its computationally cheaper surrogate-model.

The most popular choice for building surrogate models in the field of engineering design is based on taking *Gaussian processes* as a parametric family of priors along with assuming there is white noise in the observations $\epsilon_i = \mathcal{N}(0, \sigma^2)$. By definition [11, 22] Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution. Consequently it is completely specified by its mean function $m(x) = \mathbb{E}[f(x)]$ and the covariance function

$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$. The most important property of Gaussian processes for Bayesian optimization is that posterior takes also a form of Gaussian process, moreover its mean and covariance functions have analytic expressions:

- $\hat{f}_N(x) = \mathbb{E}[f(x)|D_N] = m(x) + \mathbf{k}(x)\mathbf{K}^{-1}(y_N - \mathbf{m})$,
- $\hat{\sigma}^2_N(x) = \mathbb{V}[f(x)|D_N] = k(x, x) - \mathbf{k}(x)\mathbf{K}^{-1}\mathbf{k}(x)^T$,

where $\mathbf{m} = \{m(x_i)\}_{i=1}^N$, $\mathbf{k}(x) = \{k(x, x_i) + \sigma^2\delta(x, x_i)\}_{i=1}^N$, $\mathbf{K} = \{k(x_i, x_j) + \sigma^2\delta(x_i, x_j)\}_{i,j=1}^N$
Several conditions on the family of priors that ensure convergence of the Bayesian optimization to the optimum are specified in [12], they are satisfied by a Gaussian process with a constant mean function. Therefore, the convergence is guaranteed when the principle of minimum expected risk is used, as for instance an expected deviation from the global optimum:

$$x_{i+1} = \underset{x}{\operatorname{argmin}}\ \mathbb{E}[\|f(x) - f(x_*)\||D_i] \tag{2.1}$$

Expression 2.1 is computationally expensive since it requires estimation of $f(x_*)$, thus in present work we replace this criterion with its common approximation called *Expected Improvement* [13]:

$$x_{i+1} = \underset{x}{\operatorname{argmin}}\ \mathbb{E}[\max\{0, f(x) - f(x_+)\}|D_i], \tag{2.2}$$

where $x_+ = \underset{j=1..i}{\max}\ y_i$ Expected Improvement has also been shown to converge under additional mild assumptions [14], but its main advantage is a closed-form expression, that doesn't require numerical integration:

$$\mathbb{E}[\max\{0, f(x) - f(x_+)\}|D_i] = \hat{\sigma}^2_N(x)Z\Phi(Z) + \hat{\sigma}^2_N(x)\phi(Z), \tag{2.3}$$

where $Z = \frac{\hat{f}_N(x) - f(x_+)}{\hat{\sigma}^2_N(x)}$, ϕ and Φ are PDF and CDF of the standard normal distribution respectively.

3. SHiP shield optimization

A critical component of SHiP is the muon shield, which deflects the high flux of muons produced in the target, that would represent a very serious background for the particle searches, away from the detector. The shield consists of 8 magnets and each magnet is parameterised by 7 values: length, width, etc. Because the cost of the muon shield is significant we include the weight of the muon shield as a proxy for the cost. Therefore, our aim is to find the most efficient solution at a lowest cost possible. We apply Bayesian optimization method to this task. The optimization is performed for already chosen material, thus the material balance is discussed elsewhere [15].

For the evaluation of the shield performance we use 18 million simulated events passed through the detection configuration described in [15]. In the simulation, proton fixed target collisions, inelastic neutrino interactions, and inelastic muon interactions are generated by PYTHIA8 [16], GENIE [17], and PYTHIA6 [18] respectively. Heavy flavour cascade production is also taken into account [19]. The SHiP detector response is simulated in the GEANT4 [20] framework. The simulation is done within the FAIRROOT framework [21].

The transverse (x, y) position of muons is obtained at the last tracking station by tracing them 64m downstream of the shield. For positively charged muons with $|y| < 5$m, their x-position is converted into $\sigma_{\mu^+} = \sqrt{1 - (x_\mu/\text{cm} + 300)/560}$ for $-300 < x < 260$cm, else $\sigma_{\mu^+} = 0$ [15]. σ_{μ^-} for negatively charged muons is described similarly.

Now we can introduce the complete loss function, that depends on the physical performance of the shield, its weight and the length, the later implicitly via the weight.

$$L(\Sigma, W) = (1 + \Sigma)(1 + \exp^{\frac{W - W_0}{W_0}}),$$

where $\Sigma = \sum_\mu \sigma_\mu$, W is a weight of configuration and W_0 is a weight of a baseline configuration [15] ahown in Fig.1. The weight of the baseline is about 1900 tons and Σ is approximately equal to 32. As can be seen, weight is penalized exponentially because we are not interested in the heavy regions, as we will not be able to construct such configurations due to budget constraints.

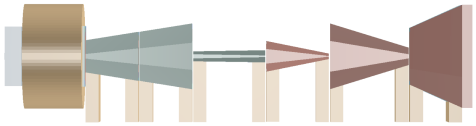

Figure 1: Baseline configuration

The following problems is addressed during the optimization:

(i) Optimization is performed in 42-dimensional space

(ii) Computation of the Σ is time consuming.

(iii) Computation of the Σ is noisy due to limited statistics in Monte - Carlo simulations.

We choose only 500k muons with the biggest momentum and discard all the low-momentum muons. It helps us to decrease the time of computations by factor of 8 times in average. We also introduce noise as a prior knowledge into Gaussian Process: in this setup GP tries to estimate the noise of the computation and incorporate it into final variance. The points are computed in batches of 100 points to optimise the usage of the available computing resources. The process of optimization is illustrated in Fig. 2

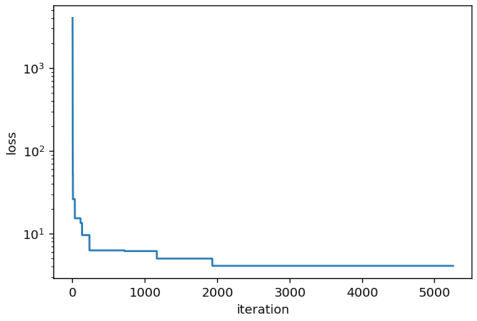

Figure 2: Evolution of the best known point

The optimization procedure is stopped after 5000 iterations. The obtained configuration is found to be lighter by 25 % than the baseline while having the same rejection capability.

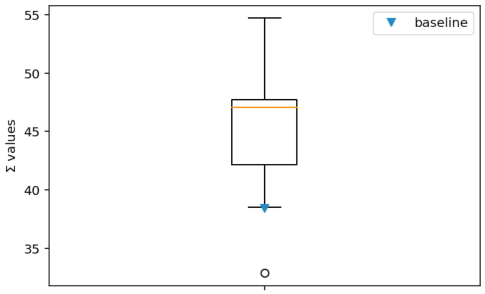

Figure 3: Behaviour of the discovered configuration

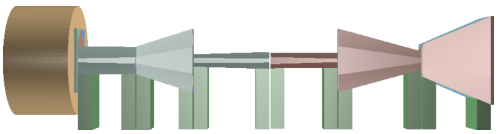

Figure 4: Discovered configuration

Fig. 3 indicates the distribution of the Σ for the discovered configuration after multiple simulations with different random seeds. We can see that the mean of the distribution is bigger than Σ for the baseline, but values are still low. Fig. 4 illustrates the new found configuration.

4. Conclusion

Bayesian optimization is a powerful tool which can be applied for the optimization of non-differentiable functions. We have demonstrated that this method can be successfully applied even to complicated optimization problems in physics. However in spite of GP can guarantee global optionality under ideal conditions, these are not satisfied for this particular problem.

Acknowledgments

References

[1] Ellis J *et. al* 2009 *Phys. Lett.* B **679** 369-375
[2] Asaka *et. al* 2005 *Phys. Lett.* B **620** 17-26
[3] Asaka *et. al* 2005 *Phys. Lett.* B **631** 151-156
[4] Bergsma F *et. al* 1986 *Phys. Lett.* B **166** 437-478
[5] Bernardi G *et. al* 1986 *Phys. Lett.* B **166** 479-483
[6] Vaitaitis A *et. al* 1999 *Phys. Rev. Lett.* **83** 4943-4946
[7] Cooper-Sarkar *et. al* 1985 *Phys. Lett.* B **160** 207-211
[8] Bonivento W *et. al* 2013 Proposal to Search for Heavy Neutral Leptons at the SPS *Preprint 1310.1762*
[9] Burnaev E and Panov M *et. al* 2015 *Proc. SLDS 2015* (Egham) 116-125
[10] Betrò B 1991 *J. of Glob. Opt.* **1** 1-14
[11] Rasmussen C E and Williams C K I 2006 *Gaussian Processes for Machine Learning* (University Press Group Limited)
[12] Mockus J 1994 *J. of Glob. Opt.* **4** 347-365
[13] Mockus J, Tiesis V and Zilinskas A 1978 *Towards Global Optimization* **2** (Elsevier Science Ltd.) 117-129
[14] Vazquez E and Bect J 2010 *J. of Stat. Plan. and Inf.* **140** 3088-3095
[15] Akmete A *et al.* 2017 *JINST* **12**
[16] Sjöstrand T, Mrenna S, and Skands P 2008 *Comput.Phys.Commun.* **178** 852-867
[17] Andreopoulos C, Bell A, Bhattacharya D, Cavanna F, Dobson J, *et al* 2010 *Nucl.Instrum.Meth.* **A614** 87-104
[18] Sjöstrand T, Mrenna S, and Skands P 2006 *JHEP* **05** 026
[19] Dijkstra H and Ruf T 2015 *SHiP-NOTE-2015-009*
[20] Agostinelli S *et al.* 2003 *Nucl.Instrum.Meth.* **A506** 250-303
[21] Al-Turany M Bertini D Karabowicz R Kresan D Malzacher P *et al.* 2012 *J.Phys.Conf.Ser.* **396**
[22] Burnaev E Panov M and Zaytsev A 2016 *Journal of Communications Technology and Electronics* **61** 6 661-671.