Towards a PIRL framework for efficient airflow diffuser design

Alfredo López
SCCH Software Competence Center Hagenberg
Softwarepark 32a, 4232 Hagenberg, Austria
alfredo.lopez@scch.at

Florian Sobieczky
SCCH Software Competence Center Hagenberg
Softwarepark 32a, 4232 Hagenberg, Austria
florian.sobieczky@scch.at

Christopher Lackner
CERBSim GmbH
Taubstummengasse 11. 1040 Vienna, Austria
clackner@cerbsim.com

Matthias Hochsteger
CERBSim GmbH
Taubstummengasse 11. 1040 Vienna, Austria
mhochsteger@cerbsim.com

Bernhard Scheichl
Institute of Fluid Mechanics and Heat Transfer
TU Wien, BA Tower/E322, 7th floor, Getreidemarkt 9, 1060 Vienna, Austria
bernhard.scheichl@tuwien.ac.at

Helmuth Sobieczky
Institute of Fluid Mechanics and Heat Transfer
TU Wien, BA Tower/E322, 7th floor, Getreidemarkt 9, 1060 Vienna, Austria
hesobi@gmail.com

Christoph Feichtinger
Windpuls Gmb
Wiener Strasse 131, 4020 Linz, Austria
c.feichtinger@windpuls.com

Abstract

This extended abstract presents a physics-informed reinforcement learning framework for optimal diffuser design to improve airflow homogeneity upstream of a heat exchanger. This approach addresses key challenges in simulation-based optimization, including high-dimensional design spaces, expensive CFD evaluations, and the lack of gradient information. Physics-based flow features related to early pressure loss occurrence and eddy formation were employed as low-cost proxies

The Third Austrian Symposium on AI, Robotics, and Vision (AIROV26).

for the target homogeneity objective. The problem is formulated as a partially observable Markov decision process in which the agent sequentially selects the geometries to be evaluated. Using an expected improvement reward function, the method adaptively balances exploration and exploitation. The approach is demonstrated on a synthetic one-dimensional example, and a two-dimensional diffuser optimization problem is presented.

1 Introduction

In recent years, there has been a growing interest in integrating reinforcement learning (RL) into optimization algorithms. Rather than adhering to a predefined set of optimization rules, the agent learns the optimization parameters and search strategy through interactions with the objective function [6, 18, 7, 16, 8]. On the other hand, the incorporation of physical information into RL has helped bridge the gap between simulated environments and real-world applications by providing a physics-based mathematical framework for addressing key RL challenges, such as exploration in high-dimensional search spaces and the design of meaningful reward functions [4, 9].

The aim of the research project [1] is to utilize physics-informed reinforcement learning (PIRL) to design an optimal channel geometry that directs the (incompressible) airflow produced by a ventilator towards a heat exchanger, with the goal of maximizing the homogeneity of the flow before it reaches the exchanger. This extended abstract formulates the PIRL optimization problem for an airflow diffuser shape design as a Partially Observable Markov Decision Process (POMDP). As each CFD run is computationally expensive (ranging from approximately 3 minutes in the present 2D example to several hours in the ongoing industrial application), each new shape iteration must be selected carefully. Within this framework, the design of new shapes is treated as a sequential decision-making process in which each CFD evaluation provides information that can either reduce the uncertainty or improve the objective function. The proposed methodology is illustrated using a synthetic one-dimensional example, and an ongoing application to diffuser shape optimization is outlined.

2 Problem formulation

2.1 Physics-informed black-box optimization

Black-box optimization (BBO) [3, 2] optimizes objective functions whose gradients and internal structures are either unavailable or too complex to handle. In this study, we address the black-box optimization problem

$$\max_{x \in X} f_{\mathrm{obj}}(x), \tag{1}$$

where $X \subset \mathbb{R}^d$ is a finite search space, and $f_{\mathrm{obj}} : \mathbb{R}^d \to \mathbb{R}$ is a continuous objective function. The derivatives and internal structure of f_{obj} are inaccessible, and each evaluation $x \mapsto f_{\mathrm{obj}}(x)$ requires a costly computer simulation. We also consider a physics-informed feature mapping $f_{\mathrm{feat}} : \mathbb{R}^d \to \mathbb{R}^p$, which extracts low-dimensional features $f_{\mathrm{feat}}(x)$ that capture physically meaningful characteristics that influence the objective function. We also assume f_{feat} is computationally inexpensive; therefore, an additional evaluation of this mapping does not constitute a significant overhead. We then define the vector-valued function

$$f(x) = \big(f_{\mathrm{obj}}(x), f_{\mathrm{feat}}(x)\big), \tag{2}$$

which jointly represents costly objective values and inexpensive physics-informed features. Our aim is to develop a computationally efficient iterative optimization strategy that, after n iterations, generates a sequence of query points $X_n := (x_t)_{t=0}^n \subseteq X$, such that X_n leads to a near-optimal design, that is, $\max_{x \in X_n} f_{\mathrm{obj}}(x) \approx \max_{x \in X} f_{\mathrm{obj}}(x)$.

2.2 POMDP

POMDPs provide a mathematical framework for RL problems in which the environment state is hidden from the agent and can only be observed indirectly [17]. A POMDP is defined as a tuple (S, A, Ω, T, O, R), where S, A, and Ω are the state, action, and observation spaces, respectively. The dynamics are specified by a transition model T, an observation model O and a reward function R.

The agent-environment interaction generates a random trajectory $s_0, a_1, o_1, r_1, s_1, a_2, o_2, r_2, s_2, \ldots$, where the initial state is drawn from an initial belief distribution $s_0 \sim b_0$. At each time step $t = 1, 2, \ldots$, the agent selects an action $a_t \sim \pi(h_t)$ according to a policy π that depends on the agent's history $h_t = (a_1, o_1, \ldots, a_{t-1}, o_{t-1})$. The environment then transitions to a new state according to the distribution $s_t \sim T(s_{t-1}, a_t)$, after which the agent receives an observation $o_t \sim O(s_t, a_t)$ and a reward $r_t = R(s_{t-1}, a_t)$. The agent then updates its history as $h_{t+1} = (h_t, a_t, o_t)$ and maintains a belief state $b_t(s) = P(s_t = s \mid h_t)$, and the process is repeated. The objective is to find a policy π that maximizes the expected discounted return $V_\pi(h) = \mathbb{E}\left[\sum_{k=t}^{t+m-1} \gamma^{k-t} r_k \,\middle|\, h_t = h\right]$, where $\gamma \in (0, 1]$ is the discount factor and $m \geq 1$ is a (possibly infinite) time horizon.

2.3 Black-box optimization as a POMDP

The optimization process is modeled as a partially observable Markov decision process (POMDP), where the latent state includes the unknown function f given in eq. (2). Because f cannot be observed directly, the agent only has access to noisy evaluations. At each time step $t = 1, 2, \ldots$, the agent selects a design point $x_t \in X$ and observes a noisy evaluation $(y_t, z_t) = f(x_t) + \varepsilon_t$. The reward r_t encourages improvements over the best objective value observed so far. More formally, the induced POMDP is defined as follows:

Initial belief $b_0 = \mathcal{GP}(0, k)$, a multi-output Gaussian process prior over the unknown vector-valued response function $f = (f_{\text{feat}}, f_{\text{obj}})$ and k is a kernel function.

State space $S = C(\mathbb{R}^{d+1}) \times X$, where $s_t = (f, x_t^*)$ consists of the latent response function f and the current best design x_t^* among the points where the objective has been evaluated.

Action space $A = X$, the finite set of candidate query points to be sampled.

Observation space $\Omega = \mathbb{R}^{p+1}$ corresponds to (possibly noisy) vector function evaluations.

Transition model The latent function remains unchanged, and the best design is updated $T(f, x_{t-1}^*, x_t) = (f, x_t^*)$, where $x_t^* = \arg\max_{x \in \{x_{t-1}^*, x_t\}} f_{\text{obj}}(x)$.

Observation model After selecting x_t, the agent observes an eventually noisy evaluation of the form $o_t = (y_t, z_t) = f(x_t) + \sigma\varepsilon_t$ with $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$.

Reward function The reward corresponds to the improvement over the previous best value, $R(f, x_{t-1}^*, x_t) = \max\big(0, f_{\text{obj}}(x_t) - f_{\text{obj}}(x_{t-1}^*)\big)$.

The following proposition shows that, under a single-step horizon and noiseless observations, our POMDP formulation coincides with standard Bayesian optimization using the expected improvement acquisition function [10, 5]. For conciseness, we omit the proof.

Proposition 1 (Bayesian Optimization)**.** *If $m = 1$, $\gamma = 1$ and $\sigma = 0$, then the optimal policy is $\pi^*(h_t) = \arg\max_{x \in X} q(x_{t-1}^*, x)$ where*

$$q(x_{t-1}^*, x) = \mathbb{E}\left[\max\big(0, f_{\text{obj}}(x) - f_{\text{obj}}(x_{t-1}^*)\big) \,\middle|\, \{(x_k, y_k, z_k)\}_{k=0}^{t-1}\right],$$

The expectation is taken under the Gaussian process posterior distribution of f_{obj} conditioned on the past evaluations $\{(x_k, y_k, z_k)\}_{k=0}^{t-1}$.

3 Experiments and analysis

In this section, we first demonstrate and evaluate the proposed method using a one-dimensional synthetic example. We then present an ongoing real-world application of diffuser shape optimization.

3.1 Illustrative 1D-example

We apply the proposed method to maximize the objective function $f_{\text{obj}}(x) = -\sin(x) - x^2 + 0.7x$ over the discrete domain X, consisting of an equidistant grid of $n = 101$ points between -1 and 2 (Figure 1). As a feature function, we consider $f_{\text{feat}}(x) = f_{\text{obj}}(x) + x$, which serves as a simple, highly correlated proxy for auxiliary information. In practical CFD settings, such features would typically not follow an explicit functional relationship with the objective but instead arise from physically correlated quantities. . To highlight the potential benefits of incorporating additional

physical information, we compared the proposed framework in two settings: without additional features ($f = f_{\text{obj}}$) and physics-informed ($f = (f_{\text{obj}}, f_{\text{feat}})$). For illustrative purposes, we assume noiseless observations ($\sigma = 0$) and consider a one-step time horizon ($m = 1$).

The process is initialized with a set of six exploratory actions $X_6 = \{-1, 0.41, 0.8, 1.01, 1.61, 1.91\}$, leading to the associated initial observations $Y_6 = \{f_{\text{obj}}(x) : x \in X_6\}$ and $Z_6 = \{f_{\text{feat}}(x) : x \in X_6\}$. The current best design is $x_6^* = 1.01$. In practice, such initial points can be selected based on expert knowledge.

Under this setting, we solve our POMDP problem using the Partially Observable Monte Carlo Planning (POMCP) method introduced in [15] and implemented in the Python library `pomdp_py` [19]. Figure 1 compares the results of the featureless (left column) and physics-informed (right column) cases. The upper row shows the objective function f_{obj} (orange dashed line), sampled objective data (crosses), sampled feature data (circles), and the expected value (blue line) and confidence interval (light blue area) of the posterior distribution $b_t(f_{\text{obj}}) = \mathbb{E}[f_{\text{obj}} \mid X_6, Y_6]$. The second row shows the action-value function $x \mapsto q(x_6^*, x)$ defined in Proposition 1. The vertical lines in the second row correspond to the next point to be queried, x_7, as determined by the algorithm, which coincides with $\arg\max_{x \in X} q(x_6^*, x)$ as stated in Proposition 1. Observe that, in the featureless case, the next queried point x_7 corresponds to a local minimizer of f_{obj}. In contrast, when additional physical information is incorporated, the selected point x_7 coincides with the global minimizer of f_{obj}.

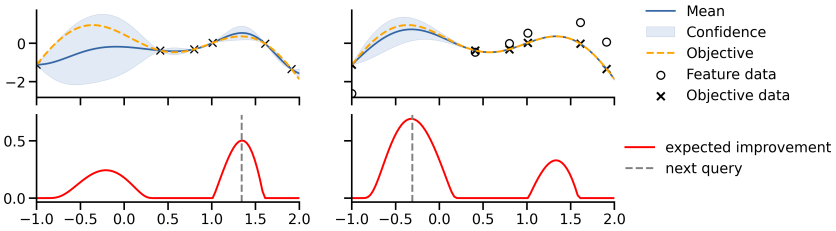


Figure 1: POMDP black-box optimization applied to a 1D example for the cases without features (left column) and physics-informed features (right column).

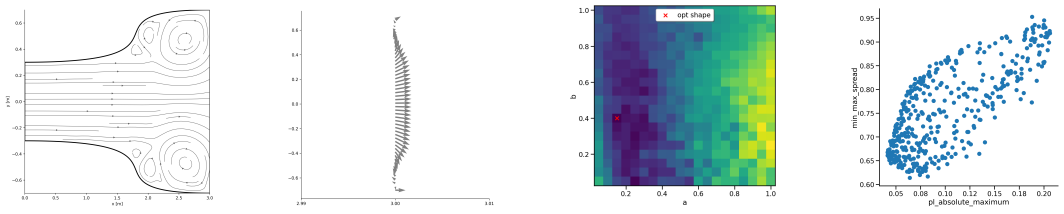
3.2 Towards diffuser shape optimization

This section outlines the application of the proposed framework to diffuser shape optimization. As this work is presented as an extended abstract, we limit the description to the main ideas and omit the detailed technical aspects.

A two-dimensional symmetric diffuser contour shape is parameterized by a vector $x \in [0, 1]^2$, which determines the position of the Breezier control point. The resulting flow field was evaluated through computational fluid dynamics (CFD) simulations based on a mass-conserving mixed stress formulation for time-dependent implicit large eddy simulations [12, 11]. Figure 2a illustrates the diffuser shape for $x = (1, 0.25)$, along with a snapshot of the corresponding velocity field. Each CFD evaluation is computationally expensive, and gradient information is unavailable, therefore this problem naturally falls within the scope of black-box optimization. We formulate the shape optimization task as in (1), where f_{obj} corresponds to a suitable homogeneity metric of the outlet velocity field, and $X \subset [0, 1]^2$ is a 20×20 regular grid of candidate designs. The heatmap in 2c shows $f_{\text{obj}}(x)$ for $x \in X$ and the optimal shape (red cross).

Modeling flow fields is challenging due to the high dimensionality of CFD data and their strong variability across geometries. Therefore, an important step in physics-informed machine learning is to extract low-dimensional, physics-based features from CFD data, guided by domain expertise [14, 13]. In our case of airflow channels, such features are selected based on their ability to capture flow regimes of enhanced mixing, which correlates with the objective of homogeneity at the outlet. More precisely, we assesed the pressure drop of the ideal pressure recovery at the outlet (estimated for a virtually steady inviscid and irrotational flow using Bernoulli's equation) and the actual one found by CFD (finite-elements) simulations, including RANS/temporal averaging. A samller pressure loss indicates a more developed turbulent regime, which enhance momentum redistribution and results in a more homogeneous RANS-averaged flow at the outlet.

In this context, the feature map f_{feat} is defined by a vector derived from the early pressure loss and vortex formation features. These features were identified by combining expert knowledge and exploratory data analysis. We obtained a set of more than ten candidate features that exhibited a strong correlation with the target and were therefore included in feature mapping. For instance, Figure 2d shows a scatter plot of $f_{\text{obj}}(x)$ against the maximum absolute value of the pressure loss associated with x for all $x \in X$. Now having all the building blocks required to apply our POMDP framework to diffuser shape optimization, we can define an appropriate covariance kernel structure for the function $f = (f_{\text{obj}}, f_{\text{feat}})$ so as to extend the computational framework from a one-dimensional search space to a multi-dimensional setting.



(a) (b) (c) (d)

Figure 2: (a) Diffuserhannel geometry and a snapshot of the velocity field for a representative shape; (b) velocity field at the diffuser outlet; (c) heatmap of the objective function over the shape search space; (d) scatter plot of the objective value versus a low-cost physical feature.

4 Conclusion and Outlook

We defined a POMDP-based black-box optimization framework that is suitable for airflow channel design optimization problems. The proposed method leads to an iterative process in which each evaluated new shape provides information by reducing uncertainty or improving the objective function, thereby fulfilling the exploration–exploitation paradigm. Our roadmap consists of the following tasks:

- Extend the one-dimensional search space case to the multidimensional setting and evaluate the method on the application described in Section 3.2.
- Extend the action space to pairs of the form $a_t = (x_t, d_t)$, where d_t is a binary decision variable that determines the evaluation fidelity: if $d_t = 0$, only the inexpensive feature map $f_{\text{feat}}(x_t)$ is evaluated, whereas if $d_t = 1$, both $f_{\text{feat}}(x_t)$ $f_{\text{obj}}(x_t)$ are evaluated.
- Apply the method to a real industrial diffuser shape optimization problem, where geometries are described by more than 30 parameters and each CFD evaluation requires several hours of computation on high-performance computing systems.

Finally, we note that the proposed framework can also be applied to optimization problems beyond shape design that may benefit from the incorporation of physics-informed features.

Acknowledgments and Disclosure of Funding

The main contributor to this research was the FFG (Austrian Applied Research Fund) project “AIrFoil” (FFG-No. 915010). The research reported in this paper has also been partly funded by the Federal Ministry for Innovation, Mobility and Infrastructure (BMIMI), the Federal Ministry for Economy, Energy and Tourism (BMWET), and the State of Upper Austria in the frame of the SCCH competence center INTEGRATE [(FFG grant no. 892418)] in the COMET - Competence Centers for Excellent Technologies Programme managed by Austrian Research Promotion Agency FFG.

References

[1] AIrFoil. Airfoil — ffg-project no. 915010 (austrian applied research foundation). https://projekte.ffg.at/projekt/5125991, 2024. FFG Projektdatenbank - Effizienzsteigerung und Lärmminderung von Luftwärmepumpen.

[2] Stéphane Alarie, Charles Audet, Aïmen E. Gheribi, Michael Kokkolaras, and Sébastien Le Digabel. Two decades of blackbox optimization applications. *EURO Journal on Computational Optimization*, 9:100011, 2021.

[3] Charles Audet and Warren Hare. *Derivative-Free and Blackbox Optimization.* Springer Series in Operations Research and Financial Engineering. Springer International Publishing, Cham, 2017.

[4] Chayan Banerjee, Kien Nguyen, Clinton Fookes, and Maziar Raissi. A survey on physics informed reinforcement learning: Review and open problems. *Expert Systems with Applications*, 287:128166, 2025.

[5] Eric Brochu, Vlad M. Cora, and Nando de Freitas. A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *ArXiv*, abs/1012.2599, 2010.

[6] Camille Castera and Peter Ochs. From learning to optimize to learning optimization algorithms, 2024.

[7] Tianlong Chen, Xiaohan Chen, Wuyang Chen, Howard Heaton, Jialin Liu, Zhangyang Wang, and Wotao Yin. Learning to optimize: A primer and a benchmark. In *AutoML Conference 2023 (Journal Track)*, 2023.

[8] Mujin Cheon, Jay H. Lee, Dong-Yeun Koh, and Calvin Tsay. EARL-BO: Reinforcement Learning for Multi-Step Lookahead, High-Dimensional Bayesian Optimization. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 10182–10198. PMLR, 2025.

[9] Thomas P. Dussauge and Chong-Kun Sung. A reinforcement learning approach to airfoil shape optimization. *Scientific Reports*, 13(1):9753, Jun 2023.

[10] Roman Garnett. *Bayesian Optimization*. Cambridge University Press, 2023.

[11] Jay Gopalakrishnan, Philip L. Lederer, and Joachim Schöberl. A mass conserving mixed stress formulation for the stokes equations. *IMA Journal of Numerical Analysis*, 40(3):1838–1874, 2020.

[12] Philip L. Lederer, Xaver Mooslechner, and Joachim Schöberl. High-order projection-based upwind method for implicit large eddy simulation. *Journal of Computational Physics*, 493: 112492, 2023.

[13] Riccardo Margheritti, Onofrio Semeraro, Maurizio Quadrio, and Giacomo Boracchi. Feature Extraction from Flow Fields: Physics-Based Clustering and Morphing with Applications. *Applied Sciences*, 15(23):12421, November 2025.

[14] Pushan Sharma, Wai Tong Chung, Bassem Akoush, and Matthias Ihme. A review of physics-informed machine learning in fluid mechanics. *Energies*, 16(5), 2023.

[15] David Silver and Joel Veness. Monte-carlo planning in large pomdps. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010.

[16] Lei Song, Chen-Xiao Gao, et al. Reinforced in-context black-box optimization. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2025*, pages 7237–7245, 2025.

[17] Matthijs T. J. Spaan. *Partially Observable Markov Decision Processes*, pages 387–414. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[18] Michael Volpp, Lukas P. Fröhlich, Kirsten Fischer, Andreas Doerr, Stefan Falkner, Frank Hutter, and Christian Daniel. Meta-learning acquisition functions for transfer learning in bayesian optimization. In *International Conference on Learning Representations (ICLR)*, 2020.

[19] Kaiyu Zheng and Stefanie Tellex. pomdp_py: A framework to build and solve pomdp problems. In *ICAPS 2020 Workshop on Planning and Robotics (PlanRob)*, 2020.